## WO9813501

Publication Title:

3' GENOMIC PROMOTER REGION AND POLYMERASE GENE MUTATIONS RESPONSIBLE FOR ATTENUATION IN VIRUSES OF THE ORDER DESIGNATED MONONEGAVIRALES

Abstract:

Abstract of WO9813501

Isolated, recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA viruses of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene are described. Vaccines are formulated comprising such viruses and a physiologically acceptable carrier. The vaccines are used for immunizing an individual to induce protection against a nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales. Data supplied from the esp@cenet database - Worldwide

Courtesy of http://v3.espacenet.com

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

## **PCT**

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/45, 15/47, 15/40, C07K 14/12, 14/115, 14/135, C12N 7/04, A61K 39/165, 39/155, C07K 14/145, 14/08

(11) International Publication Number: WO 98/13501

(43) International Publication Date: 2 April 1998 (02.04.98)

US

(21) International Application Number: PCT/US97/16718

(22) International Filing Date: 19 September 1997 (19.09.97)

27 September 1996 (27.09.96)

(71) Applicants (for all designated States except US): AMERICAN
CYANAMID COMPANY [US/US]; Five Giralda Farms,
Madison, NJ 07940 (US). THE GOVERNMENT OF THE
UNITED STATES OF AMERICA as represented by THE
DEPARTMENT OF HEALTH AND HUMAN SERVICES

DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Suite 325, 6011 Executive Boulevard, Rockville,

MD 20852 (US).

(30) Priority Data:

60/026,823

(72) Inventors; and
(75) Inventors/Applicants (for US only): UDEM, Stephen, A. [US/US]: Apartment 6F/G, 155 West 70th Street, New York, NY 10023 (US). SIDHU, Mohinderjit, S. [US/US]; 3 Aspen Court, Highland Mills, NY 10930 (US). TATEM, Joanne, M. [US/US]; 62 Douglas Drive, Towaco, NJ 07082 (US). MURPHY, Brian, R. [US/US]; 5410 Tuscarawas Road, Bethesda, MD 20816 (US). RANDOLPH, Valerie, B. [US/US]; 535 Pine Brook Road, Lincoln Park, NJ 07035 (US).

(74) Agents: GORDON, Alan, M. et al.; American Home Products Corporation, Patent Law Dept. - 2B, One Campus Drive, Parsippany, NJ 07054 (US).

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, RO, RU, SD, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ZW, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

#### Published

Without international search report and to be republished upon receipt of that report.

(54) Title: 3' GENOMIC PROMOTER REGION AND POLYMERASE GENE MUTATIONS RESPONSIBLE FOR ATTENUATION IN VIRUSES OF THE ORDER DESIGNATED MONONEGAVIRALES

### (57) Abstract

Isolated, recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA viruses of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene are described. Vaccines are formulated comprising such viruses and a physiologically acceptable carrier. The vaccines are used for immunizing an individual to induce protection against a nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales.

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

Sweden

Singapore

SE SG

LK LR

DK

EE

Denmark

Estonia

Sri Lanka

Liberia

	AL	Albania	ES	Spain	1.5	Lesotho	SI	Slovenia
	AM	Armenia	FI	Finland	LT	Lithuania	sk	Slovakia
	AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
	AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
	AZ.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
	BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Тодо
	BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
	BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
	BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkcy
	BG	Bulgaria	HU	Hungary	ML	Mali	7"	Trinidad and Tobago
	BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
	BR	Brazil	IL	Israci	MR	Mauritania	UG	Uganda
			is	Iceland	MW	Malawi	US	United States of America
	BY	Belarus	IT	Italy	MX	Mexico	UZ	Uzbekistan
	CA	Canada	JP	•	NE	Niger	VN	Viet Nam
	CF	Central African Republic	KE	Japan Kenya	NI.	Netherlands	YU	Yugoslavia
	CG	Congo		•	NO	Norway	ZW	Zimbabwe
	CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	•2**	
	CI	Côte d'Ivoire	KP	Democratic People's	Pi,	Poland		
	CM	Cameroon		Republic of Korea				
	CN	China	KR	Republic of Korea	PT	Portugal		
	CU	Cuba	KZ	Kazakstan	RO	Romania		
ľ	CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
	DE	Germany	1.1	Liechtenstein	SD	Sudan		
				Cai Lanka	CE	Sweden		

3' GENOMIC PROMOTER REGION AND POLYMERASE GENE MUTATIONS RESPONSIBLE FOR ATTENUATION IN VIRUSES OF THE ORDER DESIGNATED MONONEGAVIRALES

## 5 Field Of The Invention

10

15

This invention relates to isolated, recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA viruses of the Order designated Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene. This invention was made with Government support under a grant awarded by the Public Health Service. The Government has certain rights in the invention.

### Background Of The Invention

20 Enveloped, negative-sense, single stranded RNA viruses are uniquely organized and expressed. genomic RNA of negative-sense, single stranded viruses serves two template functions in the context of a nucleocapsid: as a template for the synthesis of 25 messenger RNAs (mRNAs) and as a template for the synthesis of the antigenome (+) strand. Negativesense, single stranded RNA viruses encode and package their own RNA dependent RNA Polymerase. Messenger RNAs are only synthesized once the virus has been uncoated 30 in the infected cell. Viral replication occurs after synthesis of the mRNAs and requires the continuous synthesis of viral proteins. The newly synthesized antigenome (+) strand serves as the template for generating further copies of the (-) strand genomic 35 RNA.

- 2 -

The polymerase complex actuates and achieves transcription and replication by engaging the cisacting signals at the 3' end of the genome, in particular, the promoter region. Viral genes are then transcribed from the genome template unidirectionally from its 3' to its 5' end. There is always less mRNA made from the downstream genes (e.g., the polymerase gene (L)) relative to their upstream neighbors (i.e., the nucleoprotein gene (N)). Therefore, there is always a gradient of mRNA abundance according to the position of the genes relative to the 3'-end of the genome.

5

10

15

20

25

30

35

Based on the revised reclassification in 1993 by the International Committee on the Taxonomy of Viruses, an Order, designated Mononegavirales, has been established. This Order contains three families of enveloped viruses with single stranded, nonsegmented RNA genomes of minus polarity (negative-sense). These families are the Paramyxoviridae, Rhabdoviridae and Filoviridae. The family Paramyxoviridae has been further divided into two subfamilies, Paramyxovirinae and Pneumovirinae. The subfamily Paramyxovirinae contains three genera, Paramyxovirus, Rubulavirus and Morbillivirus. The subfamily Pneumovirinae contains the genus Pneumovirus.

The new classification is based upon morphological criteria, the organization of the viral genome, biological activities and the sequence relationships of the proteins. The morphological distinguishing feature among enveloped viruses for the subfamily Paramyxovirinae is the size and shape of the nucleocapsids (diameter 18mm, 1mm in length, pitch of 5.5 nm), which have a left-handed helical symmetry. The biological criteria are: 1) antigenic cross-reactivity between members of a genus, and 2) the presence of neuraminidase activity in the genera Paramyxovirus,

- 3 -

Rubulavirus and its absence in genus Mopbillivirus. In addition, variations in the coding potential of the P gene are considered, as is the presence of an extra gene (SH) in Rubulaviruses.

5 Pneumoviruses can be distinguished from Paramyxovirinae morphologically because they contain narrow nucleocapsids. In addition, pneumoviruses have major differences in the number of protein-encoding cistrons (10 in pneumoviruses versus 6 in 10 Paramyxovirinae) and an attachment protein (G) that is very different from that of Paramyxovirinae. Although the paramyxoviruses and pneumoviruses have six proteins that appear to correspond in function (N, P, M, G/H/HN, F and L), only the latter two proteins exhibit 15 significant sequence relatedness between the two subfamilies. Several pneumoviral proteins lack counterparts in most of the paramyxoviruses, namely the nonstructural proteins NS1 and NS2, the small hydrophobic protein SH, and a second protein M2. 20 paramyxoviral proteins, namely C and V, lack counterparts in pneumoviruses. However, the basic genomic organization of pneumoviruses and paramyxoviruses is the same. The same is true of rhabdoviruses and filoviruses. Table 1 presents the 25 current taxonomical classification of these viruses, together with examples of each genus.

## Table 1

Classification of Nonsegmented, negative-sense, single stranded RNA Viruses of the Order Mononegavirales

Family Paramyxoviridae

Subfamily Paramyxovirinae

35

Genus Paramyxovirus

Sendai virus (mouse parainfluenza virus type 1)

- 4 -

	Human parainfluenza virus (PIV) types 1
	and 3
	Bovine parainfluenza virus (BPV) type 3
	Genus Rubulavirus
5	Simian virus 5 (SV) (Canine
	parainfluenza virus type 2)
	Mumps virus
	Newcastle disease virus (NDV) (avian
	Paramyxovirus 1)
10	Human parainfluenza virus types 2, 4a
	and 4b
	Genus Morbillivirus
	Measles virus (MV)
	Dolphin Morbillivirus
15	Canine distemper virus (CDV)
	Peste-des-petits-ruminants virus
	Phocine distemper virus
	Rinderpest virus
	Subfamily Pneumovirinae
20	Genus Pneumovirus
	Human respiratory syncytial virus (RSV)
	Bovine respiratory syncytial virus
	Pneumonia virus of mice
	Turkey rhinotracheitis virus
25	Family Rhabdoviridae
	Genus Lyssavirus
	Rabies virus
	Genus Vesiculovirus
	Vesicular stomatitis virus
30	Genus Ephemerovirus
	Bovine ephemeral fever virus
	Family Filovirdae
	Genus Filovirus

Marburg virus

- 5 -

For many of these viruses, no vaccines of any kind are available. Thus, there is a need to develop vaccines against such human and animal pathogens. Such vaccines would have to elicit a protective immune response in the recipient. The qualitative and quantitative features of such a favorable response are extrapolated from those seen in survivors of natural virus infection, who, in general, are protected from reinfection by the same or highly related viruses for some significant duration thereafter.

5

10

15

20

25

30

35

A variety of approaches can be considered in seeking to develop such vaccines, including the use of:
(1) purified individual viral protein vaccines (subunit vaccines); (2) inactivated whole virus preparations; and (3) live, attenuated viruses.

Subunit vaccines have the desirable feature of being pure, definable and relatively easily produced in abundance by various means, including recombinant DNA expression methods. To date, with the notable exception of hepatitis B surface antigen, viral subunit vaccines have generally only elicited short-lived and/or inadequate immunity, particularly in naive recipients.

Formalin inactivated whole virus preparations of polio (IPV) and hepatitis A have proven safe and efficacious. In contrast, immunization with similarly inactivated whole viruses such as respiratory syncytial virus and measles virus vaccines elicited unfavorable immune responses and/or response profiles which predisposed vaccinees to exaggerated or aberrant disease when subsequently confronted with the natural or "wild-type" virus.

Early attempts (1966) to vaccinate young children using a parenterally administered formalininactivated RSV vaccine. Unfortunately, several field

- 6 -

trials of this vaccine revealed serious adverse reactions -- the development of a severe illness with unusual features following subsequent natural infection with RSV (Bibliography entries 1,2). It has been suggested that this formalinized RSV antigen elicited an abnormal or unbalanced immune response profile, predisposing the vaccinee to RSV disease (3,4).

5

10

15

20

25

30

Thereafter, live, attenuated RSV vaccine candidates were generated by cold passage or chemical mutagenesis. These RSV strains were found to have reduced virulence in seropositive adults. Unfortunately, they proved either over or underattenuated when given to seronegative infants; in some cases, they also were found to lack genetic stability Another vaccination approach using parenteral administration of live virus was ineffective and efforts along this line were discontinued (7). Notably, these live RSV vaccines were never associated with disease enhancement as observed with the formalininactivated RSV vaccine described above. Currently, there are no RSV vaccines approved for administration to humans, although clinical trials are now in progress with cold-passaged, chemically mutagenized strains of RSV designated A2 and B-1.

Appropriately attenuated live derivatives of wild-type viruses offer a distinct advantage as vaccine candidates. As live, replicating agents, they initiate infection in recipients during which viral gene products are expressed, processed and presented in the context of the vaccinee's specific MHC class I and II molecules, eliciting humoral and cell-mediated immune responses, as well as the coordinate cytokine patterns, which parallel the protective immune profile of survivors of natural infection.

- 7 -

This favorable immune response pattern is contrasted with the delimited responses elicited by inactivated or subunit vaccines, which typically are largely restricted to the humoral immune surveillance arm. Further, the immune response profile elicited by some formalin inactivated whole virus vaccines, e.g., measles and respiratory syncytial virus vaccines developed in the 1960's, have not only failed to provide sustained protection, but in fact have led to a predisposition to aberrant, exaggerated, and even fatal illness, when the vaccine recipient later confronted the wild-type virus.

5

10

15

20

25

30

While live, attenuated viruses have highly desirable characteristics as vaccine candidates, they have proven to be difficult to develop. The crux of the difficulty lies in the need to isolate a derivative of the wild-type virus which has lost its disease-producing potential (i.e., virulence), while retaining sufficient replication competence to infect the recipient and elicit the desired immune response profile in adequate abundance.

Historically, this delicate balance between virulence and attenuation has been achieved by serial passage of a wild-type viral isolate through different host tissues or cells under varying growth conditions (such as temperature). This process presumably favors the growth of viral variants (mutants), some of which have the favorable characteristic of attenuation.

Occasionally, further attenuation is achieved through chemical mutagenesis as well.

This propagation/passage scheme typically leads to the emergence of virus derivatives which are temperature sensitive, cold-adapted and/or altered in their host range -- one or all of which are changes

- 8 -

from the wild-type, disease-causing viruses -- i.e., changes that may be associated with attenuation.

Several live virus vaccines, including those for the prevention of measles and mumps (which are paramyxoviruses), and for protection against polio and rubella (which are positive strand RNA viruses), have been generated by this approach and provide the mainstay of current childhood immunization regimens throughout the world.

5

10

15

20

25

30

Nevertheless, this means for generating attenuated live virus vaccine candidates is lengthy and, at best, unpredictable, relying largely on the selective outgrowth of those randomly occurring genomic mutants with desirable attenuation characteristics. The resulting viruses may have the desired phenotype in vitro, and even appear to be attenuated in animal models. However, all too often they remain either under- or overattenuated in the human or animal host for whom they are intended as vaccine candidates.

Even as to current vaccines in use, there is still a need for more efficacious vaccines. For example, the current measles vaccines provide reasonably good protection. However, recent measles epidemics suggest deficiencies in the efficacy of current vaccines. Despite maternal immunization, high rates of acute measles infection have occurred in children under age one, reflecting the vaccines' inability to induce anti-measles antibody levels comparable to those developed following wild-type measles infection (8,9,10). As a result, vaccine-immunized mothers are less able to provide their infants with sufficient transplacentally-derived passive antibodies to protect the newborns beyond the first few months of life.

- 9 -

Acute measles infections in previously immunized adolescents and young adults point to an additional problem. These secondary vaccine failures indicate limitations in the current vaccines' ability to induce and maintain antiviral protection that is both abundant and long-lived (11,12,13). Recently, yet another potential problem was revealed. The hemagglutinin protein of wild-type measles isolated over the past 15 years has shown a progressively increasing distance from the vaccine strains (14). This "antigenic drift" raises legitimate concerns that the vaccine strains may not contain the ideal antigenic repertoire needed to provide optimal protection. Thus, there is a need for improved vaccines.

Rational vaccine design would be assisted by a better understanding of these viruses, in particular, by the identification of the virally encoded determinants of virulence as well as those genomic changes which are responsible for attenuation.

20

25

30

35

5

10

15

### Summary Of The Invention

Accordingly, it is an object of this invention to identify those regions of the genome of the RNA viruses of the Order Mononegavirales where mutations result in attenuation of those viruses.

It is a further object of this invention to produce recombinantly-generated viruses which incorporate such attenuating mutations in their genomes.

It is still a further object of this invention to formulate vaccines containing such attenuated viruses.

These and other objects of the invention as discussed below are achieved by the generation and

- 10 -

isolation of recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA viruses of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene.

5

10

15

20

25

30

In the case of measles virus, at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 26  $(A \rightarrow T)$ , nucleotide 42  $(A \rightarrow T \text{ or } A \rightarrow C)$  and nucleotide 96 (G  $\rightarrow$  A), where these nucleotides, as well as others delineated in this application (unless stated otherwise), are presented in positive strand, antigenomic, that is, message (coding) sense, and at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 331 (isoleucine  $\rightarrow$  threonine), 1409 (alanine  $\rightarrow$  threonine), 1624 (threonine  $\rightarrow$  alanine), 1649 (arginine  $\rightarrow$ methionine), 1717 (aspartic acid → alanine), 1936 (histidine → tyrosine), 2074 (glutamine → arginine) and 2114 (arginine  $\rightarrow$  lysine).

In the case of human parainfluenza virus type 3, at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 23 (T  $\rightarrow$  C), nucleotide 24 (C  $\rightarrow$  T), nucleotide 28 (G  $\rightarrow$  T) and nucleotide 45 (T  $\rightarrow$  A), and at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 942 (tyrosine  $\rightarrow$  histidine), 992 (leucine  $\rightarrow$ 

- 11 -

phenylalanine), 1292 (leucine  $\rightarrow$  phenylalanine), and 1558 (threonine  $\rightarrow$  isoleucine).

5

10

15

20

25

30

35

In the case of human respiratory syncytial virus subgroup B, at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 4 (C  $\rightarrow$  G) and the insertion of an additional A in the stretch of A's at nucleotides 6-11, and at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 353 (arginine  $\rightarrow$  lysine), 451 (lysine  $\rightarrow$  arginine), 1229 (aspartic acid  $\rightarrow$  asparagine), 2029 (threonine  $\rightarrow$  isoleucine) and 2050 (asparagine  $\rightarrow$  aspartic acid).

In another embodiment of this invention, attenuated virus is used to prepare vaccines which elicit a protective immune response against the wild-type form of the virus.

In yet another embodiment of this invention, an isolated, positive strand, antigenomic message sense nucleic acid molecule (or an isolated, negative strand genomic sense nucleic acid molecule) having the complete viral nucleotide sequence (whether of wild-type virus or virus attenuated by non-recombinant means) is manipulated by introducing one or more of the attenuating mutations described in this application to generate an isolated, recombinantly-generated attenuated virus. This virus is then used to prepare vaccines which elicit a protective immune response against the wild-type form of the virus.

In still another embodiment of this invention, such a complete wild-type or vaccine viral nucleotide sequence is used: (1) to design PCR primers for use in a PCR assay to detect the presence of the

- 12 -

corresponding virus in a sample; or (2) to design and select peptides for use in an ELISA to detect the presence of the corresponding virus in a sample.

### Brief Description Of The Figures

5

10

15

20

25

30

35

Figure 1 depicts the passage history of the Edmonston measles virus (15). The abbreviations have the following meanings: HK - human kidney; HA - human amnion; CE(am) - chick embryo; CEF - chick embryo fibroblast; DK - dog kidney; WI-38 - human diploid cells; SK - sheep kidney; \* - plaque cloning. The number following each abbreviation represents the number of passages.

Figure 2 depicts a map of the measles virus genome showing putative cis-acting regulatory elements at and near the genome and antigenome termini. Top - a schematic map of the measles virus genome, beginning at the 3' end with 52 nucleotides of leader sequence (1) and ending at the 5' terminus with 37 nucleotides of trailer sequence (t). Gene boundaries are denoted by vertical bars; below each gene is the number of cistronic nucleotides. Bottom - an expanded schematic view of the 3' extended genomic promoter regions of genome and antigenome, showing the position and sequence of the two highly conserved domains, A and B. The intervening intergenic trinucleotide is denoted as well. Nascent 5' RNAs encompassing the A' to B' regions are presumed to contain the regulatory sequence at which the N protein encapsidation initiates.

Figure 3 depicts a genetic map of the RSV subgroup B wild-type strains designated 2B and 18537 (top portion), the intergenic sequences of those strains (middle portion) and the 68 nucleotide overlap between the M2 and L genes (bottom portion). The RSV

- 13 -

2B stain has six fewer nucleotides in the G gene, encoding two fewer amino acid residues in the G protein, as compared to the 18537 strain. The 2B strain has 145 nucleotides in the 5' trailer region, as compared to 149 nucleotides in the 18537 strain. The 2B strain has one more nucleotide in each of the NS-1, NS-2 and N genes, and one fewer nucleotide in each of the M and F genes, as compared to the 18537 strain.

5

10

15

20

25

30

35

### Detailed Description Of The Invention

Transcription and replication of negativesense, single stranded RNA viral genomes are achieved
through the enzymatic activity of a multimeric protein
acting on the ribonucleoprotein core (nucleocapsid).
Naked genomic RNA cannot serve as a template. Instead,
these genomic sequences are recognized only when they
are entirely encapsidated by the N protein into the
nucleocapsid structure. It is only in that context
that the genomic and antigenomic terminal promoter
sequences are recognized to initiate the
transcriptional or replication pathways.

All paramyxoviruses require the two viral proteins, L and P, for these polymerase pathways to proceed. The pneumoviruses, including RSV, also require the transcription elongation factor, M2, for the transcriptional pathway to proceed efficiently. Additional cofactors may also play a role, including perhaps the virus-encoded NS1 and NS2 proteins, as well as perhaps host-cell encoded proteins.

However, considerable evidence indicates that it is the L protein which performs most, if not all, the enzymatic processes associated with transcription and replication, including initiation, and termination of ribonucleotide polymerization, capping and

- 14 -

polyadenylation of mRNA transcripts, methylation and perhaps specific phosphorylation of P proteins. The L protein's central role in genomic transcription and replication is supported by its large size, sensitivity to mutations, and its catalytic level of abundance in the transcriptionally active viral complex (16).

5

10

15

20

25

30

35

These considerations led to the proposal that L proteins consist of a linear array of domains whose concatenated structure integrates discrete functions (17). Indeed, three such delimited, discrete elements within the negative-sense virus L protein have been identified based on their relatedness to defined functional domains of other well-characterized proteins. These include: (1) a putative RNA template recognition and/or phosphodiester bond formation domain; (2) an RNA binding element; and (3) an ATP binding domain. All prior studies of L proteins of nonsegmented negative-sense, single stranded RNA viruses have revealed these putative functional elements (17).

Without being bound by the following, it is reasonable to presume that these non-protein coding, promoter and other cis-acting genomic regulatory domains are important determinants of the efficiency with which transcription and replication by measles virus (MV) and other viruses of the Order Mononegavirales are actualized, in association with the L protein, and that they may therefore be virulence determinants for these viruses as well.

In summary, the invention is believed to encompass a coordinate set of changes between the cisacting regulatory signal (3' genomic promoter region) and the polymerase gene (L) which results in attenuation of the virus while retaining sufficient ability of the virus to replicate. Attenuation is

- 15 -

optimized by rational mutations of the 3' genomic promoter region and the polymerase gene, which provide the desired balance of replication efficiency: so that the virus vaccine is no longer able to produce disease, yet retains its capacity to infect the vaccinee's cells, to express sufficiently abundant gene products to elicit the full spectrum and profile of desirable immune responses, and to reproduce and disseminate sufficiently to maximize the abundance of the immune response elicited.

5

10

15

20

25

30

35

Without being bound by the following, attenuating mutations in the extended promoter (3' genomic promoter region) and in the polymerase gene are believed to affect the display of cis-acting signals and the conformation of the polymerase complex engaging these signals. For example, when encapsidated, the promoter RNA is coiled in a helical array. Changes in promoter sequence may affect the relative positions at which the conserved signals are displayed relative to one another. Specifically, the measles wild-type 3' genomic promoter region has a pyrimidine (uracil) at positions 26 and 42 (the antigenomic message sense sequences have the purine adenine). The vaccine strains have purines at those positions (the antigenomic message sense sequences have the corresponding pyrimidines; see Table 3 in Example 1 The larger purines may change the distance and/or angular display between the conserved domains of the promoter (e.g, in measles, positions 1-11 and 87-98), resulting in an altered spatial presentation of the cis-acting signals to the polymerase.

Animal studies have demonstrated a decrease in viral replication sufficient to avoid illness but adequate to elicit the desired immune response. This likely represents a decrease in transcription, a

- 16 -

decrease in gene expression of virally encoded proteins, a decrease in antisense templates and, therefore, the production of fewer new genomes. The resulting attenuated viruses are significantly less virulent than the wild-type.

The attenuating mutations described herein may be introduced into viral strains by two methods:

5

10

15

20

25

30

35

- mutagenesis during virus growth in cell cultures to which a chemical mutagen has been added, selection of virus that has been subjected to passage at suboptimal temperature in order to select temperature sensitive and/or cold adapted mutations, identification of mutant virus that produce small plaques in cell culture, and passage through heterologous hosts to select for host range mutations. These viruses are then screened for attenuation of their biological activity in an animal model. Attenuated viruses are subjected to nucleotide sequencing of their 3' genomic promoter region and polymerase genes to locate the sites of attenuating mutations. Once this has been done, method (2) is then carried out.
- attenuating mutations comprises making predetermined mutations using site-directed mutagenesis. These mutations are identified either by method (1) or by reference to closely-related viruses whose attenuating mutations are already known. One or more mutations are introduced into each of the 3' genomic promoter region and the polymerase gene. Cumulative effects of different combinations of coding and non-coding changes can also be assessed.

The mutations to the 3' genomic promoter region and polymerase gene are introduced by standard recombinant DNA methods into a DNA copy of the viral

5

10

15

20

25

30

35

genome. This may be a wild-type or a modified viral genome background (such as viruses modified by method (1)), thereby generating a new virus. Infectious clones or particles containing these attenuating mutations are generated using the cDNA "rescue" system, which has been applied to a variety of viruses, including Sendai virus (18); measles virus (19); respiratory syncytial virus (20); rabies (21); vesicular stomatitis virus (VSV) (15); and rinderpest virus (23); these references are hereby incorporated by reference. See, for measles virus rescue, published International patent application WO 97/06270, designating the United States (24); for PIV-3 rescue, U.S. provisional patent application 60/047575 (25); for RSV rescue, published International patent application WO 97/12032, designating the United States (26); these applications are hereby incorporated by reference.

Briefly, all Mononegavirales rescue systems can be summarized as follows: Each requires a cloned DNA equivalent of the entire viral genome placed between a suitable DNA-dependent RNA polymerase promoter (e.g., the T7 RNA polymerase promoter) and a self-cleaving ribozyme sequence (e.g., the hepatitis delta ribozyme) which is inserted into a propagatable bacterial plasmid. This transcription vector provides the readily manipulable DNA template from which the RNA polymerase (e.g., T7 RNA polymerase) can faithfully transcribe a single-stranded RNA copy of the viral antigenome (or genome) with the precise, or nearly precise, 5' and 3' termini. The orientation of the viral genomic DNA copy and the flanking promoter and ribozyme sequences determine whether antigenome or genome RNA equivalents are transcribed. Also required for rescue of new virus progeny are the virus-specific trans-acting proteins needed to encapsidate the naked,

- 18 -

single-stranded viral antigenome or genome RNA transcripts into functional nucleocapsid templates: the viral nucleocapsid (N or NP) protein, the polymerase-associated phosphoprotein (P) and the polymerase (L) protein. These proteins comprise the active viral RNA-dependent RNA polymerase which must engage this nucleocapsid template to achieve transcription and replication.

5

10

15

20

25

30

35

The trans-acting proteins required for measles virus rescue are the encapsidating protein N, and the polymerase complex proteins, P and L. For PIV-3, the encapsidating protein is designated NP, and the polymerase complex proteins are also referred to as P and L. For RSV, the virus-specific trans-acting proteins include N, P and L, plus an additional protein, M2, the RSV-encoded transcription elongation factor.

Typically, these viral trans-acting proteins are generated from one or more plasmid expression vectors encoding the required proteins, although some or all of the required trans-acting proteins may be produced within mammalian cells engineered to contain and express these virus-specific genes and gene products as stable transformants.

The typical (although not necessarily exclusive) circumstances for rescue include an appropriate mammallian cell milieu in which T7 polymerase is present to drive transcription of the antigenomic (or genomic) single-stranded RNA from the viral genomic cDNA-containing transcription vector. Either cotranscriptionally or shortly thereafter, this viral antigenome (or genome) RNA transcript is encapsidated into functional templates by the nucleocapsid protein and engaged by the required polymerase components produced concurrently from co-

- 19 -

transfected expression plasmids encoding the required virus-specific trans-acting proteins. These events and processes lead to the prerequisite transcription of viral mRNAs, the replication and amplification of new genomes and, thereby, the production of novel viral progeny, i.e., rescue.

5

10

15

20

25

30

35

For the rescue of rabies, VSV and Sendai, T7 polymerase is provided by recombinant vaccinia virus This system, however, requires that the rescued virus be separated from the vaccinia virus by physical or biochemical means or by repeated passaging in cells or tissues that are not a good host for poxvirus. For MV cDNA rescue, this requirement is avoided by creating a cell line that expresses T7 polymerase, as well as viral N and P proteins. Rescue is achieved by transfecting the genome expression vector and the L gene expression vector into the helper Advantages of the host-range mutant of the cell line. vaccinia virus, MVA-T7, which expresses the T7 RNA polymerase, but does not replicate in mammalian cells, are exploited to rescue RSV, Rinderpest virus and MV. After simultaneous expression of the necessary encapsidating proteins, synthetic full length antigenomic viral RNA are encapsidated, replicated and transcribed by viral polymerase proteins and replicated genomes are packaged into infectious virions. addition to such antigenomes, genome analogs have now been successfully rescued for Sendai and PIV-3 (25,27).

The rescue system thus provides a composition which comprises a transcription vector comprising an isolated nucleic acid molecule encoding a genome or antigenome of a nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating

- 20 -

mutation in the RNA polymerase gene, together with at least one expression vector which comprises at least one isolated nucleic acid molecule encoding the transacting proteins necessary for encapsidation, transcription and replication (e.g., N, P and L for measles virus; NP, P and L for PIV-3; N, P, L and M2 for RSV). Host cells are then transformed or transfected with the at least two expression vectors just described. The host cells are cultured under conditions which permit the co-expression of these vectors so as to produce the infectious attenuated virus.

5

10

15

20

25

30

35

The rescued infectious virus is then tested for its desired phenotype (temperature sensitivity, cold adaptation, plaque morphology, and transcription and replication attenuation), first by in vitro means. The mutations at the cis-acting 3' genomic promoter region are also tested using the minireplicon system where the required trans-acting encapsidation and polymerase activities are provided by wild-type or vaccine helper viruses, or by plasmids expressing the N, P and different L genes harboring gene-specific attenuating mutations (19,28).

Virus is present, challenge experiments are conducted with an appropriate animal model. Non-human primates provide the preferred animal model for the pathogenesis of human disease. These primates are first immunized with the attenuated, recombinantly-generated virus, then challenged with the wild-type form of the virus. Monkeys are infected by various routes, including but not limited to intranasal, intratracheal or subcutaneous routes of inoculation (29). Experimentally infected rhesus and cynomolgus macaques have also served as animal models for studies of

- 21 -

vaccine-induced protection against measles (30). Protection is measured by such criteria as disease signs and symptoms, survival, virus shedding and antibody titers. If the desired criteria are met, the attenuated, recombinantly-generated virus is considered a viable vaccine candidate for testing in humans. The "rescued" virus is considered to be "recombinantly-generated", as are the progeny and later generations of the virus, which also incorporate the attenuating mutations.

5

10

15

20

25

30

35

Even if a "rescued virus is underattenuated or overattenuated relative to optimum levels for vaccine use, this is information which is valuable for developing such optimum strains.

Optimally, a codon containing an attenuating point mutation may be stabilized by introducing a second or a second plus a third mutation in the codon without changing the amino acid encoded by the codon bearing only the attenuating point mutation.

Infectious virus clones containing the attenuating and stabilizing mutations are also generated using the cDNA "rescue" system described above.

Measles virus serves as a useful model for this invention, because sequence data are now available as described herein for the disease-causing wild-type virus and for the disease-preventing vaccines which have a demonstrated history of efficacy.

Measles virus was first isolated in tissue culture in 1954 (31) from an infected patient named David Edmonston. This Edmonston strain of measles became the progenitor for many live-attenuated measles vaccines including Moraten, which is the current vaccine in the United States (Attenuvax<sup>1M</sup>; Merck Sharp & Dohme, West Point, PA) and was licensed in 1968 and has proven to be efficacious.

- 22 -

Aggressive immunization programs instituted in the mid to late 1960s resulted in the precipitous drop in reported measles cases from near 700,000 in 1965 to 1500 in 1983. In parallel, other vaccine strains were also developed from the Edmonston strain (see Fig. 1), Schwarz (Institut Merieux, Lyon, France), Zagreb (Zagreb, Yugoslavia) and AIK-C (Japan). other vaccines have also proven to be efficacious and have been used extensively. An early, reactogenic, underattenuated vaccine strain (Rubeovax™: Merck Sharp & Dohme) produced measles-like illness in children and its use thus was discontinued. It, however, was further attenuated successfully to produce the Moraten vaccine strain (see Fig. 1) (32). Live measles virus vaccine provides a success story of the development of an efficacious vaccine and provides a model for understanding the molecular mechanisms of viral vaccine attenuation among nonsegmented, negative-sense, single stranded RNA viruses.

5

10

15

20

25

30

35

Because of its significance as a major cause of human morbidity and mortality, measles virus (MV) has been quite extensively studied. MV is a large, relatively spherical, enveloped particle composed of two compartments, a lipoprotein membrane and a ribonucleoprotein particle core, each having distinct biological functions (33). The virion envelope is a host cell-derived plasma membrane modified by three virus-specified proteins: The hemagglutinin (H; approximately 80 kilodaltons (kD)) and fusion (F1.2; approximately 60 kD) glycoproteins project on the virion surface and confer host cell attachment and entry capacities to the viral particle (16). Antibodies to H and/or F are considered protective since they neutralize the virus' ability to initiate infection (34,35,36). The matrix (M; approximately 37 5

10

15

20

25

30

35

kD) protein is the amphipathic protein lining the membrane's inner surface, which is thought to orchestrate virion morphogenesis and thus consummate virus reproduction (37). The virion core contains the 15,894 nucleotide long genomic RNA upon which template activity is conferred by its intimate association with approximately 2600 molecules of the approximately 60 kD nucleocapsid (N) protein (38,39,40). Loosely associated with this approximately one micron long helical ribonucleoprotein particle are enzymatic levels of the viral RNA dependent RNA polymerase (L; approximately 240 kD) which in concert with the polymerase cofactor (P; approximately 70 kD), and perhaps yet other virus-specified as well as host-encoded proteins, transcribes and replicates the MV genome sequences (41).

To date, the entire nucleotide sequences (only for the Edmonston B laboratory strain and the AIK-C vaccine strain), coding potential, and organization of the MV genome have been reported (33). The six virion structural proteins are encoded by six contiguous, non-overlapping genes which are arrayed as follows: 3'-N-P-M-F-H-L-5'. Two additional MV gene products of as yet uncertain function have also been identified. These two nonstructural proteins, known as C (approximately 20 kD) and V (approximately 45 kD), are both encoded by the P gene, the former by a second reading frame within the P mRNA; the latter by a cotranscriptionally edited P gene-derived mRNA which encodes a hybrid protein having the amino terminal sequences of P and a new zinc finger-like cysteine-rich carboxy terminal domain (16).

In addition to the sequences encoding the virus-specified proteins, the MV genome contains distinctive non-protein coding domains resembling those

5

10

15

20

25

30

35

directing the transcriptional and replicative pathways of related viruses (16,42). These regulatory signals lie at the 3' and 5' ends of the MV genome and in short internal regions spanning each intercistronic boundary. The former encode the putative promoter and/or regulatory sequence elements directing genomic transcription, genome and antigenome encapsidation, and replication. The latter signal transcription termination and polyadenylation of each monocistronic viral mRNA and then reinitiation of transcription of the next gene. In general, the MV polymerase complex appears to respond to these signals much as the RNA-dependent RNA polymerases of other non-segmented negative strand RNA viruses (16,42,43,44).

Transcription initiates at or near the 3' end of the MV genome and then proceeds in a 5' direction producing monocistronic mRNAs (40,42,45). As the polymerase traverses the MV genomic template, it encounters putative stop/start signals which, in 3' to 5' order, are: a semi-conserved transcription termination/polyadenylation signal (A/G U/C UA A/U NN A4, where N may be any of the four bases) at which each monocistronic RNA is completed; a non-transcribed intergenic trinucleotide punctuation mark (CUU; except at the H:L boundary where it is CGU); and a semiconserved start signal for transcription initiation of the next gene (AGG A/G NN C/A A A/G G A/U, where N may be any of the four bases) (45,46). Since some polymerase complexes fail to reinitiate, the abundance of each MV mRNA diminishes in parallel with the distance of the encoding gene from the genomic 3' end. This mRNA gradient directly corresponds to the relative abundance of each virus-specified protein. indicates that MV protein expression is ultimately controlled at the transcriptional level (44).

- 25 -

The 3' and 5' MV genomic termini contain non-protein coding sequences with distinct parallels to the leader and trailer RNA encoding regions of VSV (42). Nucleotides 1-55 define the region between the genomic 3' terminus and the beginning of the N gene, while 37 additional nucleotides can be found between the end of the L gene and the 5' terminus of the genome. However, unlike VSV, or even the paramyxoviruses Sendai and NDV, MV does not transcribe these terminal regions into short, unmodified (+) or (-) sense leader RNAs (47,48,49). Instead, leader readthrough transcripts, including full-length polyadenylated leader:N, leader:N:P, leader:N:P:M, and of course full-length antigenome MV RNAs are transcribed (48,49). Thus, the short leader transcript, the key operational element determining the switch from transcription to replication of the VSV single-stranded, negative polarity genome (50,51,52), seems absent in MV. This leads to consideration and exploration of alternative models for this crucial reproductive event (42).

5

10

15

20

25

30

35

Mononegavirales except the rhabdoviruses, appears to have extended its terminal regulatory domains beyond the confines of leader and trailer encoding sequences (42). For measles, these regions encompass the 107 3' genomic nucleotides (the "3' genomic promoter region", also referred to as the "extended promoter", which comprises 52 nucleotides encoding the leader region, followed by three intergenic nucleotides, and 52 nucleotides encoding the 5' untranslated region of N mRNA) and the 109 5' end nucleotides (69 encoding the 3' untranslated region of L mRNA, the intergenic trinucleotide and 37 nucleotides encoding the trailer). Within these 3' terminal approximately 100 nucleotides

- 26 -

of both the genome and antigenome are two short regions of shared nucleotide sequence: 14 of 16 nucleotides at the absolute 3' ends of the genome and antigenome are identical. Internal to those termini, an additional region of 12 nucleotides of absolute sequence identity have been located. Their position at and near the sites at which the transcription of the MV genome must initiate and replication of the antigenome must begin, suggests that these short unique sequence domains encompass an extended promoter region.

5

10

15

20

25

30

35

These discrete sequence elements may dictate alternative sites of transcription initiation -- the internal domain mandating transcription initiation at the N gene start site, and the 3' terminal domain directing antigenome production (42,48,53). addition to their regulatory role as cis-acting determinants of transcription and replication, these 3' extended genomic and antigenomic promoter regions encode the nascent 5' ends of antigenome and genome RNAs, respectively. Within these nascent RNAs reside as yet unidentified signals for N protein nucleation, another key regulatory element required for nucleocapsid template formation and consequently for amplification of transcription and replication. 2 schematically shows the location and sequence of these highly conserved, putative cis-acting regulatory domains.

Terminal non-protein coding regions similar in location, size and spacing are present in the genomes of other members of the genus Paramyxoviridae, though only 8-11 of their absolute terminal nucleotides are shared by MV (42,54). The genomic terminii of the Morbillivirus canine distemper virus (CDV) displays a greater degree of homology with its MV relative: 73% of the nucleotides of the leader and trailer sequences

- 27 -

of these two viruses are identical, including 16 of 18 at the absolute 3' termini and 17 of 18 at their 5' ends (55). No accessory internal CDV genomic domainsharing homology to that of the MV extended promoter has been found. However, there is a 20 nucleotide long stretch lying between CDV genomic nucleotides 85 and 104 and 15,587 and 15,606 in which 15 of the 20 nucleotides are complementary (Gene Bank accession number AF 14953). This indicates that CDV, like MV contains an additional region within its non-coding 3' genomic and antigenomic ends that may provide important cis-acting promoter and/or regulatory signals (55).

5

10

15

20

25

35

Additionally, the precise length of the 3'leader region (55 nucleotides) is identical among several members of the Family Paramyxoviridae (MV, CDV, PIV-3, BPV-3, SV and NDV). Further evidence for the importance of these extended, non-protein coding regions comes from analyses of a large number of distinct copy-back Defective Interfering Viruses (DIs) recently cloned from subacute sclerosing panencephalitis (SSPE) brain tissue. No DI with a stem shorter than the 95 5' terminal genomic nucleotides was This indicates that the minimal signals needed for MV DI RNA replication and encapsidation extend well beyond the 37 nucleotide long trailer sequence to encompass the additional internal putative regulatory domain (56).

As exemplified in part by measles virus, this invention is directed to the concept that important 30 virulence/attenuation determinants reside in viral genomic non-protein coding regulatory regions and in the transacting transcription/replication enzyme complex with which these cis-acting elements must interact. The cis-acting domains are found both at the 3' and 5' ends of the MV genome, flanking the six

- 28 -

contiguous genes encoding viral structural proteins; and within the MV genome as short regions encompassing internal intergenic boundaries. The former encode the putative promoter and/or regulatory sequence elements directing the vital processes of genomic transcription, genome and antigenome encapsidation, and replication. The latter signal transcription termination and polyadenylation of each monocistronic viral mRNA and then reinitiation of transcription of the next gene. The transcription/replication enzyme, RNA dependent RNA polymerase molecule can modulate transcription and/or replicative efficiency, thereby determining the abundance of cytopathic viral gene products and/or virion progeny.

5

10

15

20

25

30

proof of the concept of this invention for measles virus is obtained by first determining the nucleotide sequences of the non-coding regulatory regions (3' genomic promoter region) and the coding regions of the L gene (with predicted amino acid sequences) of the progenitor Edmonston wild-type MV isolate, together with available measles vaccine strains derived from this isolate (see Figure 1). Independent other wild-type isolates were examined for comparative purposes as well.

The nucleotide sequences (in positive strand, antigenomic, message sense) of four wild-type and five vaccine measles strains, as well as the deduced amino acid sequences of the RNA polymerase (L protein) of these measles viruses, are set forth as follows with reference to the appropriate SEQ ID NOS. contained herein:

- 29 -

	<u>Virus</u>	Nucleotide Sequence	L Protein Sequence
	Wild-Type		
	Edmonston	SEQ ID NO:1	SEQ ID NO:2
	1977	SEQ ID NO:3	SEQ ID NO:4
5	1983	SEQ ID NO:5	SEQ ID NO:6
	Montefiore	SEQ ID NO:7	SEQ ID NO:8
	<u>Vaccine</u>		
	<b>Rubeovax</b> ™	SEQ ID NO:9	SEQ ID NO:10
10	Moraten	SEQ ID NO:11	SEQ ID NO:12
	Zagreb	SEQ ID NO:13	SEQ ID NO:14
	AIK-C	SEQ ID NO:15	SEQ ID NO:16

Each measles virus genome listed above is 15,894 nucleotides in length. Translation of the L gene starts with the codon at nucleotides 9234-9236; the translation stop codon is at nucleotides 15783-15785. The translated L protein is 2,183 amino acids long.

15

20

25

30

35

Note that nucleotide 2499 of 1983 wild-type measles virus is indicated as "G" in SEQ ID NO:5. In fact, the base is actually a mixture of "G" and "C". Also note that nucleotide 2143 of Rubeovax™ vaccine virus is indicated as "T" in SEQ ID NO:9. In nine clones sequenced, this base was "T" in seven and "C" in two; thus, this base can be "T" or "C".

In addition, the Schwarz vaccine virus genome is identical to that of the Moraten vaccine virus genome (SEQ ID NO:11), except that at nucleotides 4917 and 4924, Schwarz has a "C" instead of a "T".

Nucleotide differences distinguishing the 3' genomic promoter region and nucleotide and amino acid differences distinguishing the L gene and L protein sequences of the Edmonston wild-type isolate, vaccine strains and other independently isolated wild-type

- 30 -

viruses were then compared and aligned (see Tables 3-5 in Example 1 below).

5

10

15

20

25

30

35

As shown in Table 3, there were three mutations from the 3' genomic promoter region (in antigenomic, message sense) of the progenitor wild-type MV isolate and the derivative vaccine strains: At nucleotide position 26, from "A" to "T"; at position 42, from "A" to "C" or from "A" to "T"; and in the case of Zagreb only, at position 96, from "G" to "A". In addition, the other examined wild-type isolates differed from both the progenitor wild-type isolate and the vaccine strains at position 50 by having "A" instead of "G".

The predicted amino acid sequences of the L genes of measles vaccine strains (Rubeovax™, Moraten, Schwarz, AIK-C and Zagreb) and wild-type isolates (1977, 1983 and Montefiore), differ from the progenitor strain (Edmonston) at 49 positions in the 2183 amino acid long open reading frame (see Tables 4 and 5 in Example 1 below).

These amino acid differences can be divided into four categories:

- (1) Positions where one vaccine strain differs from the progenitor, as well as from other vaccine and wild-type strains, suggesting a potential attenuation site.
- (2) Specific differences between all wildtype and all vaccine sequences; these may also constitute important attenuation sites.
- 3) Residues where chronologically newer wildtypes differ from older wild-types; which may be attributable to genetic drift.
- (4) Positions where one or more vaccine strains and/or wild-type strains have common amino acids and differ from all the other strains; these

- 31 -

changes may represent lineage-specific, potentially attenuating changes within the vaccine strains and relatedness among the wild-type isolates, respectively.

There were four category (1) changes where one vaccine differed from the other vaccines, as well as the wild-type strains. Two of these were in Moraten and Schwarz (amino acids 331 and 2114) and two were in AIK-C (1624 and 2074). These mutations are of special interest because all of these viruses are good vaccines. Thus, these positions are sites for attenuation.

5

10

15

20

25

30

35

Only one position, 1717, fits into category (2), with all wild-types having aspartic acid and all vaccines having alanine. Interestingly, this position is in one of two areas where the L genes of measles and canine distemper virus (which are otherwise highly homologous) do not show exceptional conservation. This difference makes it more likely that 1717 is a key position for an attenuating mutation in measles.

There were five positions, 149, 636, 720, 2017 and 2119, where both chronologically newer wild-types (1983 and Montefiore) differ from older wild-types (Edmonston and 1977), which therefore fit into category (3). These differences suggest genetic drift rather than denoting sites of attenuating mutations. Not included in this total are 16 positions where Montefiore (the 1989 isolate) differed from the rest (see Table 5). These could be either genetic drift (category (3)) or random change (category (4)). The remaining 23 positions are category (4), with one or more of the viruses differing from the consensus.

Three of these positions (1409, 1649, 1936) are potentially attenuating category (4) mutations. These are changes where two vaccine strains have a common change from the progenitor wild-type strain.

- 32 -

These changes may be connected with the vaccine lineage leading to the Rubeovax™ and Moraten vaccines (Figure 1).

Applicants have found that their AIK-C vaccine strain nucleotide sequence differs from the published sequence (33) at 21 positions, including one insertion and one deletion. Several of these differences result in coding changes including two in the L gene (at amino acids 1477 and 2008).

5

10

15

20

25

30

35

Thus, the additional changes accrued within the L gene sequence as the measles progenitor strain is progressively attenuated to achieve a replicative capacity optimized for live vaccine purposes appears to be constrained and delimited. Presumably, this limited tolerance in the number and location of L gene changes is imposed not only by the need to preserve the multifunctional capacities of the polymerase, but also by the preexisting 3' promoter changes with which the evolving L protein must interact to achieve transcription and replication. In other words, optimal virus attenuation requires coordinate (i.e., linked) changes in the polymerase protein and the cis-acting regulatory elements on which it acts.

The 3'-leader displays the least tolerance for change, allowing highly selected changes during the attenuation process at nucleotide position 26 (always the change of from "A" to "T"), and at position 42 (the change of from "A" to "C" or from "A" to "T") (in antigenomic, message sense). In the case of Zagreb only, there is a single further change, from "G" to "A" at position 96, which may be important when combined with Zagreb L gene-specific changes. The 3'-leader region seems to have undergone only one instance of genetic drift since 1954, with a change of "G" to "A" at position 50 (see Table 3).

- 33 -

The net change in the 3' genomic promoter region during the attenuation process is the replacement of two pyrimidines by two purines in genomic sense in all MV vaccine strains. evolution of the L gene during these attenuation processes is believed to reflect selection of subtle changes favoring reproduction of the viruses in different host cells. All the vaccine strains were grown in chick embryo (CE) or chick embryo fibroblast (CEF) cells during their attenuation process (Figure In addition, some vaccine strains have been exposed to unique host cells; i.e., Zagreb vaccine was grown in dog kidney cells and human diploid cells, while the AIK-C vaccine was adapted to sheep kidney cells. Moraten and Rubeovax™ were exclusively developed in CE and CEF.

5

10

15

20

25

30

35

Some of the lineage-specific L gene changes (position 1649 in Rubeovax<sup>™</sup>, Moraten and Schwarz vaccines and the change at position 1717 in all vaccines) represent a subset of adaptations of the L gene to the 3'-leader to modulate the transcription/replication processes for vaccine attenuation. Additionally, individual vaccine-specific changes (category (1)) may provide additional fine tune modulation of virus replication/transcription for each vaccine strain.

Based on Table 3 and the foregoing discussion, the key attenuating mutations for the MV 3' genomic promoter region are nucleotide 26 (A  $\rightarrow$  T), nucleotide 42 (A  $\rightarrow$  T or A  $\rightarrow$  C) and nucleotide 96 (G  $\rightarrow$  A) (in antigenomic, message sense).

Based on Table 4 and the foregoing discussion, the key attenuating sites for the L protein are as follows: amino acid residues 331 (isoleucine  $\rightarrow$  threonine), 1409 (alanine  $\rightarrow$  threonine), 1624

- 34 -

(threonine → alanine), 1649 (arginine → methionine), 1717 (aspartic acid → alanine), 1936 (histidine → tyrosine), 2074 (glutamine → arginine) and 2114 (arginine → lysine). It is understood that the nucleotide changes responsible for these amino acid changes are not limited to those set forth in Table 4 of Example 1 below; all changes in nucleotides which result in codons which are translated into these amino acids are within the scope of this invention.

5

10

15

20

25

30

35

Human parainfluenza virus type 3 (HPIV-3) is another nonsegmented, negative-sense, single stranded enveloped RNA virus. HPIV-3 belongs to the Family Paramyxoviridae (see Table 1). The genome of HPIV-3 is 15,462 nucleotides long and encodes six non-overlapping protein-encoding genes (57). Five of the genes encode a single virion structural protein each, which are designated NP (corresponding to the N protein of MV), M, F, HN (hemagglutinin-neuraminidase) and L. The sixth mRNA encodes the P protein, and by an overlapping 5' proximal open reading frame (ORF) encodes the C protein, and by the RNA editing mechanism, also encodes the D protein.

Like MV, HPIV-3 consists of a 3'-nonprotein coding leader region of 55 nucleotides, but unlike measles (where it is 37 nucleotides), it has a 44 nucleotide long 5'-trailer region. The polymerase transcribes the genome in a linear, sequential, startstop manner which is guided by transcription signals in the RNA template.

Attempts to develop a live attenuated HPIV-3 vaccine by passaging the wild-type virus JS strain through cell culture at sub-optimal temperature has produced promising results (7,57). Several "cold passage" (cp) mutants were isolated for evaluation from different passage levels of the JS strain. One such

- 35 -

mutant resulted from 45 serial passages and was designated cp45.

5

10

15

20

25

30

35

This virus exhibited three interesting properties: (1) cold adaptation (ca): the ability to replicate efficiently at the suboptimal temperature of 20°C; (2) temperature sensitivity (ts): inability to replicate in vitro at temperatures greater than or equal to 39°C; and (3) small plaque morphology. This mutant appeared to be a promising vaccine candidate because: (a) its ca, ts and small plaque phenotype is stable after passage in cell culture; (b) its replication is restricted in both the upper and lower respiratory tract of hamsters; and (c) it induced significant protection in hamsters against subsequent challenge with wild-type HPIV-3 (58,59).

Evaluation of this strain in the rhesus monkey showed the attenuation mutations in cp45 to be a combination of ts and non-ts mutations (60). Subsequent evaluation in chimpanzees indicated that cp45 appeared to be satisfactorily attenuated while still able to induce a high level of protection against wild-type virus challenge (61). Later preliminary clinical evaluation of cp45 in seronegative human infants and small children suggested that this candidate vaccine strain is suitably infectious and attenuated, as well as being moderately immunogenic (61).

The cp45 strain has been grown in both fetal rhesus lung (FRhL) and Vero cells as follows: The PIV-3 cp45 virus grown in FRhL cells was prepared by inoculating confluent FRhL cell monolayers in tissue culture flasks at an MOI 0.1-1.0. The infected cell cultures were fed with EMEM medium and incubated at 32°C. About seven days later, when maximal cytopathic effects (synctyia) were observed, the virus was

- 36 -

harvested by subjecting the cultures to one freeze-thaw cycle, pooling the fluids and then storing the virus at -70 °C.

The PIV-3 cp45 virus grown in Vero cells was prepared by inoculating with virus a bioreactor culture of confluent monolayers of Vero cells on microcarrier beads which was continuously stirred. The infected bioreactor culture was maintained at 30°C. The virus was harvested 4-5 days later when syncytial CPE was observed. The culture fluid containing the virus was stored at -70 °C.

5

10

15

30

35

The nucleotide sequences (in positive strand, antigenomic, message sense) of the HPIV-3 JS wild-type strain (89) and the cp45 vaccine strain grown in FRhL and Vero cells, as well as the deduced amino acid sequences of the RNA polymerase (L protein) of these HPIV-3 viruses, are set forth as follows with reference to the appropriate SEQ ID NOS. contained herein:

20	<u>Virus</u>	<u>Nucleotide Sequence</u>	L Protein Sequence
	Wild-Type		
	JS	SEQ ID NO:17	SEQ ID NO:18
	<u>Vaccine</u>		
25	FRhL cp45	SEQ ID NO:19	SEQ ID NO:20
	Vero cp45	SEQ ID NO:21	SEQ ID NO:22

Each PIV-3 virus genome listed above is 15,462 nucleotides in length. Translation of the L gene starts with the codon at nucleotides 8646-8648; the translation stop codon is at nucleotides 15345-15347. The translated L protein is 2,233 amino acids long.

As detailed in Example 2 and Table 6 therein below, based upon the differences between the wild-type

- 37 -

JS strain and the FRhL-grown cp 45 mutant vaccine strain, the key attenuating mutations for the HPIV-3 3' genomic promoter region are nucleotide 23  $(T \rightarrow C)$ , nucleotide 24  $(C \rightarrow T)$ , nucleotide 28  $(G \rightarrow T)$  and nucleotide 45  $(T \rightarrow A)$  (in antigenomic, message sense). As also detailed in Example 2 and Table 6 therein below, key attenuating sites for the L protein of HPIV-3 include the following: amino acid residues 942  $(tyrosine \rightarrow histidine)$ , 992  $(leucine \rightarrow phenylalanine)$  and 1558  $(threonine \rightarrow isoleucine)$ .

5

10

15

20

25

30

In addition, the Vero-grown cp45 mutant vaccine strain contains an additional mutation resulting from a coding change in the L gene at amino acid residue 1292 (leucine -> phenylalanine).

It is understood that the nucleotide changes responsible for these amino acid changes are not limited to those set forth in Example 2 below; all changes in nucleotides which result in codons which are translated into these amino acids are within the scope of this invention.

Human respiratory syncytial virus (RSV) is yet another nonsegmented, negative-sense, single stranded enveloped RNA virus. RSV belongs to the Subfamily Pneumovirinae and the genus Pneumovirus (see Table 1).

Two major subgroups of human RSV, designated A and B, have been identified based on reactivities of the F and G surface glycoproteins with monoclonal antibodies (62). More recently, the A and B lineages of RSV strains have been confirmed by sequence analysis (63,64). Bovine, ovine, and caprine strains of this virus have also been isolated. The host specificity of the virus is most clearly associated with the G attachment protein, which is highly divergent between

- 38 -

the human and the bovine/ovine strains (65,66), and may be influenced, at least in part, by receptor binding.

5

10

15

20

25

30

RSV is the primary cause of serious viral pneumonia and bronchiolitis in infants and young Serious disease, i.e., lower respiratory children. tract disease (LRD), is most prevalent in infants less than six months of age. It most commonly occurs in the nonimmune infant's first exposure to RSV. RSV additionally is associated with asthma and hyperreactive airways and it is a significant cause of mortality in "high risk" children with bronchopulmonary dysplasia and congenital heart disease (CHD). It is also one of the common viral respiratory infections predisposing to otitis media in children. In adults, RSV generally presents as uncomplicated upper respiratory illness; however, in the elderly it rivals influenza as a predisposing factor in the development of serious LRD, particularly bacterial bronchitis and pneumonia. Disease is always confined to the respiratory tract, except in the severely immunocompromised, where dissemination to other organs can occur. Virus is spread to others by fomites contaminated with virus-containing respiratory secretions, and infection initiates through the nasal, oral, or conjunctival mucosa.

RSV disease is seasonal and virus is usually isolated only in the winter months, e.g., from November to April in northern latitudes. The virus is ubiquitous, and over 90% of children have been infected at least once by 2 years of age. Multiple strains cocirculate. There is no direct evidence of antigenic drift (such as that seen with influenza A viruses), but sequence studies demonstrating accumulation of amino acid changes in the hypervariable regions of the G

- 39 -

protein and SH proteins suggest that immune pressure may drive virus evolution.

In mouse and cotton rat models, both the F and G proteins of RSV elicit neutralizing antibodies and immunization with these proteins alone provides longterm protection against reinfection (67,68).

5

10

15

20

25

30

35

In humans, complete immunity to RSV does not develop and reinfections occur throughout life (69,70); however, there is evidence that immune factors will protect against severe disease. A decrease in severity of disease is associated with two or more prior infections and there is evidence that children infected with one of the two major RSV subgroups may be somewhat protected against reinfection with the homologous subgroup (71), observations which suggest that a live attenuated virus vaccine may provide protection sufficient to prevent serious morbidity and mortality. Infection with RSV elicits both antibody and cell mediated immunity. Serum neutralizing antibody to the F and G proteins has been associated, in some studies, with protection from LRD, although reduction in upper respiratory disease (URD) has not been demonstrated. High levels of serum antibody in infants is associated with protection against LRD, and adminstration of intravenous immunoglobulin with high RSV neutralizing antibody titers has been shown to protect against severe disease in high risk children (70,72,73). The role of local immunity, and nasal antibody in particular, is being investigated.

The RSV virion consists of a ribonucleoprotein core contained within a lipoprotein envelope. The virions of pneumoviruses are similar in size and shape to those of all other paramyxoviruses. When visualized by negative staining and electron microscopy, virions are irregular in shape and range in

- 40 -

diameter from 150-300 nm (74). The nucleocapsid of this virus is a symmetrical helix similar to that of other paramyxoviruses, except that the helical diameter is 12-15 nm rather than 18nm. The envelope consists of a lipid bilayer that is derived from the host membrane and contains virally coded transmembrane surface glycoproteins. The viral glycoproteins mediate attachment and penetration and are organized separately into virion spikes. All members of paramyxovirus subfamily have hemagglutinating activity, but this function is not a defining feature for pneumoviruses, being absent in RSV but present in PVM (75). Neuraminidase activity is present in members of the genera Paramyxovirus, Rubulavirus, and is absent in Morbillivirus and Pneumovirus of mice (PVM) (75).

5

10

15

20

25

30

35

RSV possesses two subgroups, designated A and The wild-type RSV (strain 2B) genome is a single strand of negative-sense RNA of 15,218 nucleotides (SEQ ID NO:23) that are transcribed into ten major subgenomic mRNAs. Each of the ten mRNAs encodes a major polypeptide chain: Three are transmembrane surface proteins (G, F and SH); three are the proteins associated with genomic RNA to form the viral nucleocapsid (N, P and L); two are nonstructural proteins (NS1 and NS2) which accumulate in the infected cells but are also present in the virion in trace amounts and may play a role in regulating transcription and replication; one is the nonglycosylated virion matrix protein (M); and the last is M2, another nonglycosylated protein recently shown to be an RSVspecified transcription elongation factor (see Figure 3). These ten viral proteins account for nearly all of the viral coding capacity.

The viral genome is encapsidated with the major nucleocapsid protein (N), and is associated with

5

20

25

30

35

the phosphoprotein (P), and the large (L) polymerase protein. These three proteins have been shown to be necessary and sufficient for directing RNA replication of cDNA encoded RSV minigenomes (76). Further studies have shown that for transcription to proceed with full processing, the M2 protein (ORF 1) is required (74). When the M2 protein is missing, truncated transcripts predominate, and rescue of the full length genome does not occur (74).

Both the M (matrix protein) and the M2
proteins are internal virion-associated proteins that
are not present in the nucleocapsid structure. By
analogy with other nonsegmented negative-stranded RNA

viruses, the M protein is thought to render the

nucleocapsid transcriptionally inactive before
packaging and to mediate its association with the viral

envelope. The NS1 and NS2 proteins have only been detected in very small amounts in purified virions, and at this time are considered non-structural. Their

functions are uncertain, though they may be regulators of transcription and replication. Three transmembrane surface glycoproteins are present in virions: G, F, and SH. G and F (fusion) are envelope glycoproteins that are known to mediate attachment and penetration of the virus into the host cell. In addition, these

glycoproteins represent major independent immunogens (77). The function of the SH protein is unknown, although a recent report has implicated its involvement in the fusion function of the virus (78).

The genomes of two wild-type RSV subgroup B strains (2B and 18537) have now been sequenced in their entirety (see SEQ ID NOS:23 and 25, discussed below).

Genomic RNA is neither capped nor polyadenylated (79).

In both the virion and intracellularly, genomic RNA is tightly associated with the N protein.

5

10

15

20

25

30

- 42 -

The 3' end of the genomic RNA consists of a 44-nucleotide extragenic leader region that is presumed to contain the major viral promoter (Fig. 3). genomic promoter region is followed by ten viral genes in the order 3'-NS1-NS2-N-P-M-SH-G-F-M2-L-5' (Fig. 3). The L gene is followed by a 145-149 nucleotide extragenic trailer region (see Figure 3). Each gene begins with a conserved nine-nucleotide gene start signal 3'-GGGGCAAAU (except for the ten-nucleotide gene start signal of the L gene, which is 3'-GGGACAAAAU; differences underlined). For each gene, transcription begins at the first nucleotide of the signal. Each gene terminates with a semi-conserved 12-14 nucleotide gene end  $(3'-A G U/G U/A ANNN U/A A_{1.5})$  (where N can be any of the four bases) that directs transcription termination and polyadenylation (Fig. 3). The first nine genes are non-overlapping and are separated by intergenic regions that range in size from 3 to 56 nucleotides for RSV B strains (Fig. 3). The intergenic regions do not contain any conserved motifs or any obvious features of secondary structure and have been shown to have no influence on the preceding and succeeding gene expression in a minreplicon system (Fig. 3). The last two RSV genes overlap by 68 nucleotides (Fig. 3). The gene-start signal of the L gene is located inside of, rather than after, the M2 gene. This 68 nucleotide overlap sequence encodes the last 68 nucleotides of the M2 mRNA (exclusive of the Poly-A tail), as well as the first 68 nucleotides of the L mRNA.

Ten different species of subgenomic polyadenylated mRNAs and a number of polycistronic polyadenylated read-through transcripts are the products of genomic transcription (74). Transcriptional mapping studies using UV light mediated

- 43 -

genomic inactivation showed that RSV genes are transcribed in their 3' to 5' order from a single promoter near the 3' end (80). Thus, RSV synthesis appears to follow the single entry, sequential transcription model proposed for all Mononegavirales (16,81). According to this model, the polymerase (L) contacts genomic RNA in the nucleocapsid form at the 3' genomic promoter region and begins transcription at the first nucleotide. RSV mRNAs are co-linear copies of the genes, with no evidence of mRNA editing or splicing.

5

10

15

20

25

30

35

Sequence analysis of intracellular RSV mRNAs showed that synthesis of each transcript begins at the first nucleotide of the gene start signal (74). end of the mRNAs are capped with the structure m7G(5')ppp(5')Gp (where the underlined G is the first template nucleotide of the mRNA) and the mRNAs are polyadenylated at their 3' ends (82). Both of these modifications are thought to be made cotranscriptionally by the viral polymerase. regions of the RSV 3' genomic promoter have been found to be important as cis acting elements (83). regions are the first ten nucleotides (presumably acting as a promoter), nucleotides 21-25, and the gene start signal located at nucleotides 45-53 (83). Unlike other Paramyxovirinae, such as measles, Sendai and PIV-3, the remainder of the leader and non-coding region of NS1 gene of RSV was found to be highly tolerant of insertions, deletions and substitutions (83).

Additionally, by saturation mutagenesis
(wherein each base is replaced independently by each of
the other three bases and compared for translation and
replication efficiencies) within the first 12
nucleotides of the 3' genomic promoter region, a Utract located at nucleotides 6-10 was shown to be

- 44 -

highly inhibitory to substitutions (83). In contrast, the first five nucleotides were relatively tolerant of a number of substitutions and two of them at position four were up-regulatory mutations, resulting in a four-to 20-fold increase in RSV-CAT RNA replication and transcription. Using a bi-cistronic minireplicon system, gene-start and gene-end motifs were shown to be signals for mRNA synthesis and appear to be self contained and largely independent of the nature of adjoining sequence (84).

5

10

15

20

25

30

35

The L gene start signal lies 68 nucleotides upstream of the M2 gene-end signal, resulting in gene overlap (Fig. 3) (74). The presence of the M2 gene-end signal within the L gene results in a high frequency of premature termination of L gene transcripts. length L mRNA is much less abundant and is made when the polymerase fails to recognize the M2 gene-end motif. This results in much lower transcription of L mRNA. The gene overlap seems incompatible with a model of linear sequential transcription. It is not known whether the polymerase that exits the M2 gene jumps backward to the L gene-start signal or whether there is a second, internal promoter for L gene transcription It is also possible that the L gene is accessible by a small fraction of polymerases that fail to start transcription at the M2 gene-start signal and slide down the M2 gene to the L gene-start signal.

The relative abundance of each RSV mRNA decreases with the distance of its gene from the promoter, presumably due to polymerase fall-off during sequential transcription (80). Gene overlap is a second mechanism that reduces the synthesis of full length L mRNA. Also, certain mRNAs have features that might reduce the efficiency of translation. The initiation codon for SH mRNA is in a suboptimal Kozak

- 45 -

sequence context, while the G ORF begins at the second methionyl codon in the mRNA.

5

10

15

20

25

RSV RNA replication is thought (74) to follow the model proposed from studies with vesicular stomatitis virus and Sendai virus (16,81). involves a switch from the stop-start mode of mRNA synthesis to an antiterminator read-through mode. This results in synthesis of positive sense replicationintermediate (RI) RNA that is an exact complementary copy of genomic RNA. This serves in turn as the template for the synthesis of progeny genomes. The mechanism involved in the switch to the antiterminator mode is proposed to involve cotranscriptional encapsidation of the nascent RNA by N protein (16,81). RNA replication in RSV like other nonsegmented negative-strand RNA viruses is dependent on ongoing protein synthesis (85). Predicted RI RNA has been detected for the standard virus as well as RSV-CAT minigenome (74,85). RI RNA was 10-20 fold less abundant intracellularly than was the progeny genome both for the standard and the minigenome system. nucleotide sequences (in positive strand, antigenomic, message sense) of various wild-type, vaccine and revertant RSV strains, as well as the deduced amino acid sequences of the RNA polymerase (L protein) of these RSV viruses, are set forth as follows with reference to the appropriate SEQ ID NOS. contained herein:

- 46 -

	<u>Virus</u>	Nucleotide	Sequence	L Prote	in Sequence
	Wild-Type				
	2B	SEQ ID	NO:23	SEQ	ID NO:24
	18537	SEQ ID	NO:25	SEQ	ID NO:26
5					
	<u>Vaccine</u>				
	2B33F	SEQ ID	NO:27	SEQ	ID NO:28
	2B20L	SEQ ID	NO:29	SEQ	ID NO:30
10	Revertant				
	2B33F TS(+)	SEQ ID	NO:31	SEQ	ID NO:32
	2B20L TS(+)	SEQ ID	NO:33	SEQ	ID NO:34

Each RSV virus genome encodes an L protein that is 2,166 amino acids long. Genome length and other nucleotide information is as follows:

:

30

35

As detailed in Example 3 (especially Tables 7 and 8) below, the key attenuating mutations for the RSV subgroup B 3' genomic promoter region are nucleotide 4 (C  $\rightarrow$  G), and the insertion of an additional A in the stretch of A's at nucleotides 6-11 (in antigenomic

- 47 -

message sense). As also detailed in Example 3 below, the key potentially attenuating sites for the L protein of RSV are as follows: amino acid residues 353 (arginine  $\rightarrow$  lysine), 451 (lysine  $\rightarrow$  arginine), 1229 (aspartic acid  $\rightarrow$  asparagine), 2029 (threonine  $\rightarrow$  isoleucine) and 2050 (asparagine  $\rightarrow$  aspartic acid). It is understood that the nucleotide changes responsible for these amino acid changes are not limited to those set forth in Example 3 below; all changes in nucleotides which result in codons which are translated into these amino acids are within the scope of this invention.

5

10

15

20

25

30

35

The attenuated viruses of this invention exhibit a substantial reduction of virulence compared to wild-type viruses which infect human and animal hosts. The extent of attenuation is such that symptoms of infection will not arise in most immunized individuals, but the virus will retain sufficient replication competence to be infectious in and elicit the desired immune response profile in the vaccinee.

The attenuated viruses of this invention may be used to formulate a vaccine. To do so, the attenuated virus is adjusted to an appropriate concentration and formulated with any suitable vaccine adjuvant, diluent or carrier. Physiologically acceptable media may be used as carriers. These include, but are not limited to: an appropriate isotonic medium, phosphate buffered saline and the like. Suitable adjuvants include, but are not limited to MPL<sup>IM</sup> (3-O-deacylated monophosphoryl lipid A; RIBI ImmunoChem Research, Inc., Hamilton, MT) and IL-12 (Genetics Institute, Cambridge, MA).

In one embodiment of this invention, the formulation including the attenuated virus is intended for use as a vaccine. The attenuated virus may be mixed

- 48 -

with cryoprotective additives or stabilizers such as proteins (e.g., albumin, gelatin), sugars (e.g., sucrose, lactose, sorbitol), amino acids (e.g., sodium glutamate), saline, or other protective agents. This mixture is maintained in a liquid state, or is then dessicated or lyophilized for transport and storage and mixed with water immediately prior to administration.

5

10

15

20

25

30

35

Formulations comprising the attenuated viruses of this invention are useful to immunize a human or animal subject to induce protection against infection by the wild-type counterpart of the attenuated virus. Thus, this invention further provides a method of immunizing a subject to induce protection against infection by an RNA virus of the Order Mononegavirales by administering to the subject an effective immunizing amount of a vaccine formulation incorporating an attenuated version of that virus as described hereinabove.

A sufficient amount of the vaccine in an appropriate number of doses must be administered to the subject to elicit an immune response. Persons skilled in the art will readily be able to determine such amounts and dosages. Administration may be by any conventional effective form, such as intranasally, parenterally, orally, or topically applied to any mucosal surface such as intranasal, oral, eye, vaginal or rectal surface, such as by an aerosol spray. The preferred means of administration is by intranasal administration.

In another embodiment of this invention, an isolated nucleic acid molecule having the complete viral nucleotide sequence of either the wild-type viruses or vaccine viruses described herein is used to generate oligonucleotide probes (from either positive strand antigenomic message sense or negative strand

- 49 -

complementary genomic sense) and to express peptides (from positive strand antigenomic message sense only), which are used to detect the presence of those wild-type virus and/or vaccine strains in samples of body fluids and tissues. The nucleotide sequences are used to design highly specific and sensitive diagnostic tests to detect the presence of the virus in a sample.

5

10

15

20

25

30

35

Polymerase chain reaction (PCR) primers are synthesized with sequences based on the viral wild-type or vaccine sequences described herein. The test sample is subjected to reverse transcription of RNA, followed by PCR amplification of selected cDNA regions corresponding to the nucleotide sequence described herein which have nucleotides which are distinct for a defined strain of virus. Amplified PCR products are identified on gels and their specificity confirmed by hybridization with specific nucleotide probes.

ELISA tests are used to detect the presence of antigens of the wild-type or vaccine viral strains. Peptides are designed and selected to contain one or more distinct residues based on the wild-type or vaccine sequences described herein. These peptides are then coupled to a hapten (e.g., keyhole limpet hemocyanin (KLH) and used to immunize animals (e.g., rabbits) for the production of monospecific polyclonal antibody. A selection of these polyclonal antibodies, or a combination of polyclonal and monoclonal antibodies can then be used in a "capture ELISA" to detect antigens produced by those viruses.

Samples of the Moraten measles virus vaccine strain were deposited by Applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., under the provisions of the Budapest Treaty for the Deposit of Microorganisms for the Purposes of Patent Procedures

- 50 -

("Budapest Treaty") and have been assigned ATCC accession number VR2587. Samples of the HPIV-3 virus Vero-grown cp45 vaccine strain were deposited by Applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., under the provisions of the Budapest Treaty and have been assigned ATCC accession number VR2588. Samples of the 2B wild-type RSV virus were deposited by Applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., under the provisions of the Budapest Treaty and have been assigned ATCC accession number VR2586.

5

10

15

20

25

Given these three deposited strains and the sequence information for these and other strains provided herein, one can use site-directed mutagenesis and rescue techniques described above to introduce mutations (or restore a wild-type genotype) of all the strains described herein, as well as taking these strains and making additional mutations from the panel of mutations set forth in Tables 3, 4 and 6-8 below.

In order that this invention may be better understood, the following examples are set forth. The examples are for the purpose of illustration only and are not to be construed as limiting the scope of the invention.

## Examples

Standard molecular biology techniques are utilized according to the protocols described in Sambrook et al. (86).

- 51 -

## Example 1 Measles

Moraten MV vaccine virus was grown once, 5 directly from the Attenuvax™ vaccine vial (Lot #0716B), the Schwarz vaccine virus was grown once (Lot 96G04/M179 G41D), while the Zagreb and Rubeovax™ vaccine viruses were each grown twice in the Vero cells before RNAs were made for sequence analysis. My 10 wildtype isolate Montefiore (56) was passed 5-6 times in Vero cells before extraction of RNA materials and similarly, MV wildtype isolates 1977, 1983 (14) were grown 5-7 times before extracting materials for analysis. Edmonston wild-type isolate received from 15 Dr. J. Beeler (CBER) (see Fig. 1) was the original Edmonston isolate already passaged seven times in human kidney cells and three times in Vero cells before receipt and further passaged once in Vero cells before using for sequence analysis.

RNA was prepared by infecting Vero cells at a multiplicity of infection (m.o.i.) of 0.1 to 1.0 and allowed to reach maximum cytopathology before being harvested. Total RNA from measles virus-infected cells was extracted using Trizol™ reagent (Gibco-BRL).

25

30

The total RNA isolated from Vero cell passage material was amplified by the Reverse Transcriptase-PCR (Perkin-Elmer/Cetus) procedure using measles (Edmonston B strain (19)) specific primer pairs spanning the 3' and 5' promoter regions and the L gene of the viral genome. Table 2 presents these primer sequences. The primers of SEQ ID NOS:35-54, 74, 77 and 78 are in antigenomic message sense. The primers of SEQ ID NOS:55-73, 75, 76 and 79 are in genomic negative-sense.

- 52 -

Table 2
Primers for PCR and Sequencing MV L Genes
and Genomic Termini

5	9047CATATCACTCACTCTGGGATGGAG <sub>9070</sub>	(SEQ ID NO:35)
	9371TCAGAACATCAAGCACCGCC9390	(SEQ ID NO:36)
	<sub>9741</sub> ACAGTCAAGACTGAGATGAG <sub>9760</sub>	(SEQ ID NO:37)
	10001AAGAGTCAGATACATGTGGA <sub>10020</sub>	(SEQ ID NO:38)
	10351ACATGAATCAGCCTAAAGTC	(SEQ ID NO:39)
10	10674CCGAAAGAGTTCCTGCGTTACGACC <sub>10698</sub>	(SEQ ID NO:40)
	11083 CAGTCCACACAAGTACCAGG11102	(SEQ ID NO:41)
	11461GTCAGAAGCTGTGGACCATC11480	(SEQ ID NO:42)
	11841AATATTGCTACAACAATGGC <sub>11860</sub>	(SEQ ID NO:43)
	12196ACTCTTCATTCCTAGACTGG12215	(SEQ ID NO:44)
15	12542GTCCAATTATGACTATGAAC	(SEQ ID NO:45)
	12891AGAACAGACATGAAGCTTGC	(SEQ ID NO:46)
	13232CCAACAAGGAATGCTTCTAG13251	(SEQ ID NO:47)
	13551ACAGCACTATCTATGATTGACCTGG13575	(SEQ ID NO:48)
	13930GCAACATGGTTTACACATGC	(SEQ ID NO:49)
20	14280AGATTGAGAGTTGATCCAGG <sub>14299</sub>	(SEQ ID NO:50)
	<sub>14629</sub> AGGAGATACTTAAACTAAGC <sub>14648</sub>	(SEQ ID NO:51)
	14981 TAAGCTTATGCCTTTCAGCG <sub>15000</sub>	(SEQ ID NO:52)
	15337TTAACGGACCTAAGCTGTGC	(SEQ ID NO:53)
	15671GAAACAGATTATTATGACGG <sub>15690</sub>	(SEQ ID NO:54)
25		
	9290CGGGCTATCTAGGTGAACTTCAGG9267	(SEQ ID NO:55)
	<sub>9500</sub> ATTTGGATATGGAATATGAG <sub>9481</sub>	(SEQ ID NO:56)
	9840ACTCAACTGAACTACCAGTG <sub>9821</sub>	(SEQ ID NO:57)
	10181AAGAACATCATGTATTTCAG <sub>10162</sub>	(SEQ ID NO:58)
30	1054,TTATCAACGCACTGCTCATG <sub>10530</sub>	(SEQ ID NO:59)
	10919ATTTTCAGCAATCACTTGGCATGCC <sub>10895</sub>	(SEQ ID NO:60)
	11280GCCTCTGTGCAAACAAGCTG11261	(SEQ ID NO:61)
	11638 TCTCTAGTTACTCTAGCAGC 11619	(SEQ ID NO:62)
	$_{12010}\mathtt{AGGTCGTTGTTTGTGAGGAG}_{11991}$	(SEQ ID NO:63)
35	12361TCGTCCTTCTTTACTGTC <sub>12342</sub>	(SEQ ID NO:64)

- 53 -

	12689 CCGTCCTCGAGCTAGCCTCG12670	(SEQ ID NO:65)
	13052CTCCTCCAGGCTCACATTGG	(SEQ ID NO:66)
	13420GGGTTGGTACATAGCTCTGC13401	(SEQ ID NO:67)
	13767 CACCCATCTGATATTTCCCTGATGG	(SEQ ID NO:68)
5	14099TGGTTGACAGTACAAATCTG14080	(SEQ ID NO:69)
	14460CTGAAATGGGAAGATTGTGC14441	(SEQ ID NO:70)
	14820AGCAATCTACACTGCCTACC14801	(SEQ ID NO:71)
	15180 TCACAGATGATTCAATTATC	(SEQ ID NO:72)
	15530GATCCTAGATATAAGTTCTC15511	(SEQ ID NO:73)
10		
	1ACCAAACAAAGTTGGGTAAGG21	(SEQ ID NO:74)
	$GGGGGATCC_{\mathtt{100}}\mathtt{ATCCCTAATCCTGCTCTTGTCCC}_{\mathtt{78}}$	(SEQ ID NO:75)
	200 GATTCCTCTGATGGCTCCAC <sub>181</sub>	(SEQ ID NO:76)
	15721TAACAGTCAAGGAGACCAAAG <sub>15741</sub>	(SEQ ID NO:77)
15	$\tt GGGAAGCTT_{15801}AACCCTAATCCTGCCCTAGGTGG_{15823}$	(SEQ ID NO:78)
	15894ACCAGACAAAGCTGGGAATAGA	(SEQ ID NO:79)

Overlapping PCR fragments of the complete viral genome were directly sequenced without cloning to achieve the consensus sequence, by the dideoxy terminator cycle sequencing method using both strands (ABI PRISM 377 sequencer and ABI PRISM sequencing Kit). To determine the sequence at the absolute termini, a ligation procedure described previously was used (55).

20

25

30

To test this hypothesis, the nucleotide sequences were determined for the non-protein coding regulatory regions and the L gene of the progenitor Edmonston wild-type MV isolate, for the available vaccine strains derived from this isolate, as well as for other wild-type strains. Nucleotide (in antigenomic, message sense) and amino acid differences were then compared and aligned as set forth in Tables 3-5 (differences are in italics):

- 54 -

Table 3

Differences in MV 3' Genomic Promoter Region

Nucleotide Sequence

	Nuc	leot	ide	number:
<u>Virus</u>	<u>26</u>	<u>42</u>	<u>50</u>	<u>96</u>
Edmonston w-t	A	A	G	G
Vaccines:				
Rubeovax™	T	C	G	G
Moraten	T	C	G	G
Schwarz	T	C	G	G
Zagreb	T	T	G	A
AIK-C	T	C	G	G
Wild-Types:				
1977	A	A	A	G
1983	A	A	A	G
Montefiore	A	A	A	G

- 55 -

Table 4
Differences in MV L Nucleotides and Amino Acids
Between Edmonston Wild-Type and Vaccine Strains

	<u>331</u>	1409	1624	1649	<u>1717</u>	1887	<u>1936</u>	2074	2114
Edmonston w-t	ATT ACT	GCA ACA	ACC GCC	AGG ATG	GAT GCT	AAC GAC	CAT TAT	CAA CGA	AGA AAA
Edmonston w-t Rubeovax™ vac.	I I	A A	T T	R M	D A	N D	H H	Q O	R R
Moraten vac.	T	A	T	M	A	D	H	Q	K
Schwarz vac.	T	A	T	M	A	D	H	Q	K
Zagreb vac.	I	T	T	R	A	N	H	Q	R
AIK-C vac.	I	T	A	R	' A	N	Y	R	R

Table 5

		Dif	Differences	ces i	in MV	L Nucleotides	leoti		and Amino Acids	ino A	cide
				Ве	Between		Wild-Type	Strains	ins		
	81	122	149	252	331	441	447	200	513	570	613
Edmonston w-t	೮೮೮	GAT	GIT	ACA	ATT	AAA	AAA	GAT	GTG	AAA	TAC
Mutation	ACC	AAT	ATT	GCA	GIT	AGA	AGA	AAT	ATG	AAT	CAC
Edmonston w-t	4	A	>	H	н	×	ĸ	А	>	×	×
1977 W-t	Ø	Z	>	Н	Λ	×	×	Д	E	×	×
	H	Α	I	₽	н	×	×	×	¥	Z	Ħ
Montefiore w-t	æ	А	I	¥	н	æ	R	Д	X	×	×
	618	621	623	626	628	632	636	637	641	645	650
Edmonston w-t	GTC	AGT	AGG	AGA	GCA	ATA	CAA	GTA	GAC	GAT	ATG
Mutation	GCC	AAT	AAG	AAA	GAA	GTA	CAT	ATA	AAT	AAT	ATA
Edmonston w-t	>	Ø	ø	œ	4	н	Ø	>	Ω	Ω	E
1977 w-t	4	×	ĸ	æ	K	H	Ø	I	Ω	×	Σ
1983 w-t	>	ശ	ĸ	œ	Ø	н	H	>	Ω	Ω	Σ
Montefiore w-t	>	Ø	æ	×	βq	>	H	>	×	А	I

Table 5 (continued)

					1	)	יייייייייייייייייייייייייייייייייייייי	ייי פער			
		Di	Differences	ces :	in MV	L Nu	L Nucleotides		and Au	and Amino Acids	cids
				Ř	Between		Wild-Type Strains	Stra	ains		
	652	720	723	794	914	970	1044	1294	1569	1705	1745
Edmonston w-t	GCT	ATC	TAT	CGG	CGG	ပ္ပပ္ပ	GGA	AGC	GTT	ATC	AAT
Mutation	ACC	GIC	$\mathbf{I}GC$	TGG	SS	TCA	AGA	ACC	ATT	GTC	AGT
Edmonston w-t	4	н	₩	×	24	æ	O	တ	>	H	z
1977 w-t	æ	н	v	X	Ø	æ	ប	ຜ	>	н	z
1983 w-t	4	>	บ	×	æ	Ø	O	H	H	H	×
Montefiore w-t	E H	^	υ	<b>¤</b>	æ	4	æ	Ø	>	۸	Ø
	1860	1865	1936	2007	2013	2017	2030	2096	2119	2165	
Edmonston w-t	GTA	TIC	CAT	GAC	GAT	ACT	AAT	ATA	AAG	GIC	
Mutation	ATA	TAC	TAT	වුල්	GGT	ATT	AGT	GTA	රිහිට	ATC	
Edmonston w-t	>	<u>[24</u>	Ħ	Q	Д	H	Z	н	×	>	
1977 w-t	>	₩	Ħ	А	Q	H	z	н	×	>	
1983 w-t	>	Çe,	⋈	Ω	Ö	I	z	H	Ωţ	H	
Montefiore w-t	I	ľΣų	Ħ	O	Ω	I	Ŋ	۸	ĸ	>	

- 58 -

## Example 2 PIV-3

A comparison of sequences (in antigenomic message sense) of the parental wild-type JS strain of PIV-3 virus and the FRhL-grown and Vero-grown forms of the cp45 mutant are set forth in Table 6. Where a codon change does not result in an amino acid change, Table 6 states "none", followed by the name of the unchanged amino acid.

Sequence Comparison of Vero- and FRhL-grown cp45 & JS strains Table 6

Gene Region	Nucleotide	JS	FRhL	Vero	Codon Change	Amino Acid Change
	Position		cp45	cp45		(number in L)
3' leader	23	۲	บ	ບ		
	24	ບ	H	H		
	28	២	H	H		
	45	H	Æ	Ø		
NP UTR	62	æ	H	H		
NP coding	397	H	บ	U	1	Val → Ala
	1275	H	U	U	TCT → GCT	Set 4 Tes
P coding	2080	E	U	U	1	none/Asn
M coding	4347	ບ	A	K	CCC + ACC	14 T C10
F coding	5536	ບ	H	H	AAC → AAT	none/Asn
	6329	Æ	O	O	1	Ile → Val
	6419	U	ď	4	1	111 1 0 1 V
HN coding	6847	H	บ	บ	1	none/Glv
	7956	H	บ	U	1	e la ← lev
L coding	9323	H	υ	υ		none/Tyr (226)
	9971	K	U	Ö	1	none/Glu (442)
	11469	H	U	U	1	$Tvr \rightarrow His (942)$
	11621	U	H	H	1	Leu → Phe (992)
	12521	Ø	4	*1	1	
	12581	ບ	Н	H	TIC \ TIL	)
	13318	ບ	H	H	1	
# mutations			20	21		

- 60 -

Sequence analysis of the parental wild-type JS strain of PIV-3 virus and the FRhL-grown cp45 mutant showed that the latter contained 20 nucleotide changes. Four changes were in the noncoding 3'-leader region at nucleotide positions 23 (T  $\rightarrow$ C), 24 (C  $\rightarrow$ T), 28 (G  $\rightarrow$ T) and 45 (T  $\rightarrow$  A) (in antigenomic, message sense). When considered in the genomic, negative sense, the change at position 28 from the smaller pyrimidine ("C") to the larger purine ("A") may change the size of the region flanked by the conserved regions of the 3' genomic promoter region, resulting in an altered spatial presentation of the cis-acting signals to the polymerase.

5

10

15

20

25

30

35

Nine changes were coding changes in the NP, M, F, HN and L genes. The other seven changes were non-coding or silent changes in the NP, P, F, HN and L genes or the NP untranslated region (UTR). The cp45 mutant has been demonstrated to have poor transcription activity at non-permissive temperatures due to its ts phenotype (87). This ts phenotype has now been mapped to the viral L gene (88). Because the cp45 virus has been shown to function normally with regard to mutations in the HN and F glycoproteins (87), this supports the implication that mutations in the 3'-leader and L gene contributed to the attenuating phenotype of this virus.

Thus, the four 3' leader specific changes in FRhL-grown cp45 and the three coding changes in the L gene at amino acid positions 942 (Tyr  $\rightarrow$  His), 992 (Leu  $\rightarrow$  Phe) and 1558 (Thr  $\rightarrow$  Ile) contributed significantly to the attenuation phenotype of the candidate cp45 vaccine strain.

Furthermore, the Vero-grown cp45 mutant vaccine strain contains an additional mutation resulting from a coding change in the L gene (marked

- 61 -

with an asterisk in Table 6) at amino acid residue 1292 (leucine  $\rightarrow$  phenylalanine).

The first two amino acid changes in the L protein (at positions 942 and 992) map to one of the highly conserved areas among all Paramyxovirus L genes. The fourth amino acid change (at position 1558) maps to the area joining two conserved blocks corresponding to the change at amino acid 1717 in the MV vaccine strains.

5

10

15

20

25

The published literature (89) sets forth only 18 changes between the antigenomic message sense sequences of the JS and FRhL-grown cp45 strains. Sixteen of these changes were found by applicants.

The published literature did not report four changes found by applicants: in the 3' leader at nucleotide 45 (T  $\rightarrow$  A), in the NP UTR at nucleotide 62 (A  $\rightarrow$  T), or the changes in amino acids in the NP protein resulting from the changes at nucleotide 397 (T  $\rightarrow$  C), leading to the amino acid change (Val  $\rightarrow$  Ala) and nucleotide 1275 (T  $\rightarrow$  G), leading to the amino acid change (Ser  $\rightarrow$  Ala) (nucleotide changes in antigenomic, message sense). Nor did the published literature report the additional potentially attenuating mutation in the L protein found by applicants in the Vero-grown cp45 strain resulting from the change at nucleotide 12521 (A  $\rightarrow$  T), leading to the change in amino acid 1292 (Leu  $\rightarrow$  Phe).

- 62 -

## Example 3 RSV Subgroup B

The temperature-sensitive (ts) phenotype is strongly associated with attenuation in vivo; in addition, some non-ts mutations may also be attenuating. Identification of ts and non-ts attenuating mutations was achieved by sequence analysis and evaluation of ts, cold-adapted (ca), and in vivo growth phenotypes of RSV mutants and revertants.

5

10

15

20

25

30

The genomes of the following five RSV 2B strains have now been completely sequenced: 2B parent, 2B33F, one revertant designated 2B33F TS(+), 2B20L and one revertant designated 2B20L TS(+). The 2B33F and 2B20L strains are ts and ca and are described in U.S. Serial No. 08/059,444 (90), which is hereby incorporated by reference. After identifying regions where mutations in 2B33F and 2B20L are located, nine additional isolates of 2B33F "revertants" obtained following in vitro passaging at 39°C and in vivo passaging in African Green Monkeys or chimpanzees, and nine additional isolates of 2B20L "revertants" obtained following in vitro passaging at 39°C have been sequenced in those regions. The ts, ca, and attenuation phenotypes of many of these revertants have now been characterized and assessed. Correlations between phenotype ts, vaccine attenuation and sequence changes have been identified.

A summary of results is presented in Tables 7-12.

- 63 -

Table 7
Sequence comparison between RSV 2B and 2B33F strains

	Nucl. pos.†	Nuc	leotide	changes	
Gene/ region	3' end of vRNA	RSV 2B	RSV 2B33F	RSV 2B33F TS(+), 5a revertant	Amino acid changes
Genomic Promoter	4 6	C -	G extra A	G extra A	non-coding non-coding
М	4175 4199	T T	C C	c c	non-coding non-coding
SH	4329 4409 4420 4442 4454 4484	T T T T	טטטטטט	0 0 0 0	Phe-Leu (10) none Ile (36) Ile-Thr (40) none His (47) none Cys (51) none Tyr (61)
	4497 4505 4525 4526 4542 4561 4575	T T T T T	000000	000000	Stop-Gln (66) none Ser (68) Ile-Thr (75) Ile-Thr (75) Stop-Gln (81) Leu-Pro (87) Trp-Arg (92)
L	4598 9559 9853* 12186 14587 15071	T G A G C	C A G A T G	C A A T G	none Thr (99)  Arg-Lys (353)  Lys-Arg (451)*  Asp-Asn (1229)  Thr-Ile (2029)  non-coding

for 2B33F and 2B33F TS(+), nucl. pos. numbers are one larger than for 2B for M, SH & L genes

<sup>\*</sup> At pos. 9853, the Lys-Arg change has reverted back to Lys in the 2B33F TS(+) strain

- 64 -

Table 8
Sequence comparison between RSV 2B and 2B20L strains

	Nucl. pos.t	Nuc	leotide d	changes	
Gene/ region	3' end of vRNA	RSV 2B	RSV 2B20L	RSV 2B20L TS(+), R1 revertant	Amino acid changes
Genomic Promoter	4 6	C -	G extra A	G extra A	non-coding* non-coding*
L	8963 13347 14587 14649 14650	C A C A	T A T G A	T G T G	none Thr (154) Asn-Asp (1616) Thr-Ile(2029)* Asn-Asp (2050) Asn-Asp-Val (2050)**

- for 2B20L and 2B20L TS(+), nucl. pos. numbers are one larger than for 2B for L gene
- Mutation is common in 2B33F and 2B20L strains
- \*\* At pos. 14650, the mutation suppresses the ts phenotype in 2B20L TS(+) revertant

Table 9 RSV 2B, ts and Revertant Strains

Sample	Source	In Vitro Phenotype ts ca	enotype ca	I) Cott	In Vivo Growth* Cotton Rat	Growth*	AGM
		39/32°C KOP plaque morph	20/32°C Yield	Nasal turbinates	Lungs	Nasal Wash	Bronchial Lavage
RSV 2B	Wild-type Parent Strain	0.7 (WI)	0.0001	5.5* 3.9 <sup>b</sup> (4/4)	5.8 <sup>a</sup> 5.2 <sup>b</sup> (4/4)	5.8*	4.7"
l l	ca, ts mutant isolated from 2B cold-passaged x 33	0.00007 (sp/int/wt)	0.04	<1.6* <1.9* (1/4)	<1.5* <1.2* (0/4)	3.0*	<0.9° (0/4)
RSV 2B33F - 5a TS(+)	28337 spinner passage, plaque picked at 39°C	0.5 (WT)	0.03	≤1.7° (1/4)	3.5*	4.2	4.0*
RSV 2B33F - 4a TS(+)	2833F spinner passage, plaque picked at 39°C	0.7 (WT)	0.01	≤1.7° (3/4)	3.8*	<u>B</u>	6X
3F - 3b	age,	0.5 (WT)	0.04	≤2.5* (3/4)	2.9*	<del>S</del>	£
ади рр2	2B33F-infected AGM #A2,d7 nasal wash plaque picked at 32°C	0.3 (sp, int)	0.00002	≤2.0 <sup>b</sup> (1/4)	1.6 <sup>b</sup> (4/4)	Š.	g.

Table 9 (continued)
RSV 2B, ts and Revertant Strains

Sample	Source	In Vitro Phenotype ts	enotype	Ir	In Vivo Growth* Cotton Rat	rowth*	AGM
		39/32°C EOP plaque morph	20/32°C Yield	Nasal turbinates	Lungs	Nasal Wash	Bronchial Lavage
AGM pp4	2B33F-infected AGM #A2,d7 nasal wash plaque picked at 32°C	0.1 (sp,int)	0.008	<1.6 <sup>b</sup> (0/4)	1.2 <sup>b</sup> (4/4)	<del>S</del>	CX.
agm pp6	2B33F-infected AGM #A4,d12 nasal wash plaque picked at 32°C	0.000004 (wt)	<0.00005	<1.5 <sup>b</sup> (1/4)	<1.1 <sup>b</sup> (0/4)	Ð	ξ.
AGM pp7	2B33F-infected AGM #A4,d12 nasal wash plaque picked at 32°C	0.000004 (sp/int/wt)	0.007	<1.4 <sup>b</sup> (1/4)	<1.0 <sup>b</sup> (0/4)	Ą	QN QN
Chimp ppla	2B33F-infected Chimp #1552, d4 tracheal lavage plaque picked at 32°C	0.5 (WT)	CN CN	МD	QN	QN QN	Q.
Chimp pp3A	2B33F-infected Chimp #1560, d6 tracheal lavage plaque picked at 32°C	0.7 (WI)	QN.	2.4° (4/4)	≤3.0¢ (3/4)	ę.	QX
Chimp pp5A	2B33F-infected Chimp #1563, d10 nasal swab plaque picked at 32°C	0.7 (WT)	QX	<2.3° (3/4)	3.0° (4/4)	£	S .

Table 9 (continued)
RSV 2B, ts and Revertant Strains

Sample	Source	In Vitro Phenotype ts ca	henotype ca	I	In Vivo Growth* Cotton Rat	Growth*	AGM
		39/32°C BOP plague morph	20/32°C Yield	Nasal turbinates	Lungs	Nasal Wash	Bronchial
RSV 2B20L	ca, ts mutent isolated from 2B cold-passaged x 20	0.0002 (int/wt)	0.02	<1.9 <sup>4</sup> (0/4)	<1.3 <sup>4</sup> (0/4)	<0.7 <sup>¢</sup> (0/2)	<0.7 <sup>¢</sup> (0/2)
RSV 2B20L R1 TS(+)	2B20L spinner passage, plaque picked at 39°C	0.6 (WT)	QN.	2.3° (4/4)	3.5°	<u>g</u>	GX GX
RSV 2B20L R2 TS(+)	2820L spinner passage, plaque picked at 39°C	0.6 (WT)	CX.	≤2.5° (3/4)	2.7° (4/4)	£	GX GX
RSV 2B20L R9 TS(+)	2B20L spinner passage, plaque picked at 39°C	0.8 (WT)	Ð	≤2.2° (3/4)	4.0°	Ą	QX
RSV 2B20L R10 TS(+)	2B20L spinner passage, plaque picked at 39°C	0.7 (WI)	S	2.6°	3.2° (4/4)	Ð	<b>Q</b>

BD = small plaque size int = intermediate plaque size
 Dose = 10<sup>6.3</sup> PFU IN
 Dose = 10<sup>6.0</sup> PFU IN+IT \* In Vivo growth measured in  $\log_{10}$  mean virus titer (# infected/# total) WT = wild-type plaque size IN bose = 10<sup>3.6</sup> PFU IN IN bose = 10<sup>6.6</sup> PFU IN+IT ND = not done WT Dose = 10<sup>6.7</sup> PFU IN Dose = 10<sup>5.9</sup> PFU IN

Table 10 2B33F Revertants

	ts (+)	Inv	itro		AGM				Chimp	
	5a	4a	3b	pp2	pp4	pp6	pp7	1A	3 <b>A</b>	5A
base no. †										<u> </u>
м										
4176,4200	s	S	S	s	S	s	s	s	s	s
SH										
14 bases*	s	8	S	S	s	s	s	s	s	s
L										
9560	s	s	S	s	S	S	S	s	S	s
9854	2B	2B	2B	2B	3	s	S	ND	2B	2B
12187	s	s	S	S	s	s	s	s	S	S
14588	s	S	S	S	s	s	s	ND	S	s
15072	S	s	S	s	S	s	S	s	S	s
Phenotype										
ts	2B	2B	2B	r	r	S	s	2B	2B	2B
ca	s	S	S	2B	S	2B	s	ND	ND	ND
Attenuated	r	r	r	(r)	(r)	s	s	ND	r	r

t These 2B33F revertant base nos. are one larger than for 2B for M, SH and L genes

<sup>\*</sup> bases 4330,4410,4421,4443,4455,4485,4498,4506,4526,4527,4543, 4562,4576,4599

S = same base as 2B33F

<sup>2</sup>B = reversion to 2B base or complete reversion in phenotype

r = moderate reversion in phenotype

<sup>(</sup>r) = slight reversion in phenotype

ND = not done

Table 11
2B20L Revertants

				TS (+)	In vi	tro I	solat	8.8		
base no. †	R1	R2	R3A	R4A	R5A	R6A	R7A	R8A	R9A	R10A
L			•		<del></del>			· · · · · · · · · · · · · · · · · · ·	<del></del>	
8964	s	s	S	S	s	s	S	S	s	s
13348	C*	s	ND	S	s	ND	s	S	s	s
14588	s	s	S	S	s	s	S	s	S	s
14650	S	S	2B	s	2B	2B	S	S	2B	2B
14651	A*	A*	S	A*	S	S	A*	A*	s	S
Phenotype										
ts	2B	2B	ND	ND	ND	ND	ND	ND	2B	2B
Attenuated	F	r	ND	ND	ND	ND	ND	ND	r	r

t These 2B20L revertant base nos. are one larger than for 2B for L genes

S = same base as 2B20L

<sup>2</sup>B = reversion to 2B base

r = moderate reversion in phenotype

<sup>\* =</sup> base change, different from 2B or 2B20L

ND = not done

- 70 -

Table 12
RSV 2B, ts and Revertant Strains: Phenotype Summary

Virus Isolate	Source	In Vitro Phenotype		In Vivo Attenuation		
	en comment	t <i>s</i>	Ca	Cotton Rat	AGM	
RSV 2B	Wild-type Parent Strain	-	-	-	-	
RSV 2B33F	ca, ts mutant isolated from 2B, cold-passaged x 33	++++	++	++++	+++	
RSV 2B33F - 5a TS(+)	2B33F spinner passage plaque picked at 39°C	-	++	++	+	
RSV 2B33F - 4a TS(+)	2B33F spinner passage plaque picked at 39°C	-	++	++	ND	
RSV 2B33F - 3b TS(+)	2B33F spinner passage plaque picked at 39°C	-	++	++	ND	
AGM pp2	2B33F-infected AGM A2, d7 nasal wash plaque picked at 32°C	+	-	+++	ND	
AGM pp4	2B33F-infected AGM A2, d7 nasal wash plaque picked at 32°C	+	++	+++	ND	
AGM pp6	2B33F-infected AGM A4, d12 nasal wash plaque picked at 32°C	++++	-	++++	ND	
AGM pp7	2B33F-infected AGM A4, dl2 nasal wash plaque picked at 32°C	++++	++	++++	ND	
Chimp pplA	2B33F-infected chimp #1552, d4 tracheal lavage, plaque picked at 32°C	-	ND	ND	ND	
Chimp pp3A	2B33F-infected chimp #1560, d6 tracheal lavage, plaque picked at 32°C	-	ND	++	ND	
Chimp pp5A	2B33F-infected chimp #1563, d10 tracheal lavage, plaque picked at 32°C	-	ND	++	ND	

- 71 -

Table 12 (continued)
RSV 2B, ts and Revertant Strains: Phenotype Summary

Virus Isolate	Source	In Vitro Phenotype		In Vivo Attenuation	
		ts	ca	Cotton Rat	AGM
RSV 2B20L	ca, ts mutant isolated from 2B, cold-passaged x 20	++++	++	++++	++++
RSV 2B20L R1 TS(+)	2B20L spinner passage plaque picked at 39°C	-	ND	++	ND
RSV 2B20L R2 TS(+)	2B20L spinner passage plaque picked at 39°C	-	ND	++	ND
RSV 2B20L R9 TS(+)	2B20L spinner passage plaque picked at 39°C	-	ND	. ++	ND
RSV 2B20L R10 TS(+)	2B20L spinner passage plaque picked at 39°C	•	ND	++	ND

## ND = not done

<sup>- =</sup> wild-type phenotype, i.e., not temperature sensitive, not cold adapted, not attenuated

<sup>+</sup> to ++++ = increasing levels of temperature sensitivity, coldadaptation or attenuation

PCT/US97/16718

WO 98/13501

- 72 -

Several significant observations can be drawn from these data:

As shown in Tables 7 (for 2B33F) and 8 (for 2B2OL), there are relatively few sequence changes identified in the two mutant strains: RSV 2B33F differs from parental RSV 2B by two changes at the 3' genomic promoter region, two changes at the non-coding 5'-end of the M gene, and four coding changes plus one non-coding (poly(A) motif) change in the RNA dependent RNA polymerase coding L gene. In addition, 14 changes mapped to the SH gene alone. RSV 2B20L differs from its RSV 2B parent only at seven nucleotide positions, of which three are common with 2B33F virus, including two changes at the 3' genomic promoter and one coding 15 change in the L gene. Two additional unique changes of 2B20L virus mapped to the coding region of the L gene. Potentially attenuating mutations at the non-coding 3' genomic promoter region and the RNA dependent RNA polymerase gene have been identified. 20

5

10

- Two ts mutations can be identified in the L gene of the attenuated virus strains 2B33F and 2B20L:
- In 2B33F, a mutation at nucleotide position 25 (i) 9853 (A  $\rightarrow$  G) leading to a coding change in L protein at amino acid 451 (Lys  $\rightarrow$  Arg) is clearly associated with the ts and attenuation phenotypes. Reversion at this site alone in the 2B33F TS(+) 5a strain is responsible for complete restoration of growth at 39°C 30 (Table 9) and partial reversion in attenuation in This association with the ts and attenuation phenotypes was also supported by partial sequence analyses of six additional "full TS revertants" (designated 4a, 3b, pp2, 3A, 5a, 5A) isolated from cell 35

- 73 -

culture and from chimps, in which only the nucleotide 9853 mutation reverted (Tables 10-12) (note that one AGM (African Green Monkey) isolate which reverted at 9853 only partially reverted in ts phenotype). This amino acid 451 mutation (Lys  $\rightarrow$  Arg) is amenable to stabilization in cDNA infectious clone constructs, by inserting a second mutation to stabilize the codon, thereby lessening the likelihood that it will revert back to Lys.

10

15

20

25

5

- (ii)In 2B20L, a mutation at base 14,649 (A  $\rightarrow$  G) leading to a coding change in the L protein (amino acid position 2,050, Asn  $\rightarrow$  Asp) appears to be associated with the ts and attenuation phenotypes. This aspartic acid at the amino acid 2050 invariably reverts back (Asp → Asn) in TS(+) revertants or changes to a different amino acid (Asp  $\rightarrow$  Val) by nucleotide substitution at position 14,650 (A  $\rightarrow$  T) (Tables 8, The above observation is based on complete sequence analysis on the TS(+) revertant R1 and partial sequence of several additional TS(+) revertants (R2, R4A, R7A, R8A) at selected regions (Table 11). additional mutation is seen in the R1 revertant at nucleotide postion 13,347 (amino acid 1616, Asn → Asp) associated with the above reversion. However, the effect of this mutation on the ts phenotype is not known; the L gene of other revertants has not been sequenced completely.
- 30 c. Three base changes are common to 2B33F and 2B20L strains of virus:
  - (i) A change at position 14,587 (C  $\rightarrow$  T) with a corresponding change (Thr  $\rightarrow$  Ile) at amino acid 2029 is

- 74 -

present in both 2B33F and 2B20L (Tables 7,8). This nucleotide "T" substitution was found to be present in 10% of the population of the progenitor RSV2B strain and may have been preferred during the attenuation process. No wildtype base "C" was found in the 2B33F and 2B20L virus.

5

10

15

20

25

30

35

(ii) Two mutations are seen in the 2B33F and 2B20L 3' genomic promoter region: nucleotide 4 ( $C \rightarrow G$ ) and the insertion of an extra A in the stretch of A's at positions 6-11 (in antigenomic, message sense). When the sequences of selected TS(+) revertants were analyzed, these mutations were seen to have been retained in the 2B33F TS(+)5a (Table 7) and the 2B20L TS(+)R1 (Table 8) revertants. These non-coding, cisacting mutations remained associated with partial viral attenuation.

Expression using the minireplicon RSV-CAT system for the analysis of these cis-acting changes has shown the 3' genomic promoter nucleotide 4 ( $C \rightarrow G$ ) change to be an upregulation of transcription/replication in this in vitro system when the 2B progenitor virus or either of the 2B33F or 2B33F TS(+) provided helper L gene functions (the N, P and M2 genes are identical in these viruses).

Complementation analysis of the 2B33F 3' genomic promoter and the helper functions provided by the progenitor RSV2B virus or the 2B33F and 2B33F TS(+) viruses by this RSV-CAT minireplicon system has also been conducted. All three viruses supported both the 2B and 2B33F 3' genomic promoter mediated transcription/replication functions. However, the 2B33F and 2B33F TS(+) viruses preferred their 2B33F 3' genomic promoters. This analysis clearly shows coevolution of 3' genomic promoter changes during the

- 75 -

vaccine attenuation process, along with the RNA dependent RNA polymerase gene. Reversion of ts phenotype in the 2B33F mutant 5a by reversion of the single L protein amino acid 451 (Arg → Lys) by sequence analysis was clearly demonstrated by support of transcription/replication functions of RSV-CAT minireplicon at 37°C. The 2B33F virus did not provide helper functions to the RSV-CAT minireplicon (with 2B or 2B33F 3' genomic promoters) at 37°C.

10

15

20

25

5

d. A biased hypermutation of SH seen in 2B33F is present in all 2B33F revertants, regardless of phenotype, and is not seen in 2B20L, which is ts, ca, and attenuated. Thus, there are no data at this time that associate this mutation with any biological phenotype.

Another wild-type RSV designated 18537 was also sequenced and compared to the sequence of the wild-type RSV 2B strain. With one exception, at all the critical residues described above, the two wild-type strains were identical. For 2B, the codon ACA at nucleotides 14586-14588 encodes a Thr at amino acid 2029 of the L protein, while for 18537, the codon ATT at nucleotides 14593-14595 encodes an Ile at amino acid 2029 (the L gene start codon is at nucleotides 8509-8511 in 18537, compared to 8502-8504 in 2B).

#### Example 4

## PCR Assay to Detect Measles Virus

30

35

A 21 year old patient was admitted to a hospital with a three week history of progressive non-productive cough, shortness of breath, and fever. His symptoms failed to improve following treatment with clarithromycin for seven days or after a similar course

- 76 -

of treatment with atovaquone. Concomitant complaints of right upper quadrant abdominal pain proved recalciltrant to omeprazole and antacids. Relevant past medical history included Factor VIII deficiency and HIV infection diagnosed 3-4 years prior to this hospital admission. One year earlier, he had received a booster immunization of measles-mumps-rubella (MMR) vaccine as required for college enrollment.

5

10

15

20

25

30

35

Bronchoalveolar lavage and transbronchial biopsies performed two days after admission to the hospital demonstrated reactive hyperplasia and alveolar lining cell desquamation with minimal chronic inflammation. No microorganisms were revealed by Gram, methenamine silver, or PAS stains. CT scans of the chest showed multiple, ill-defined, confluent nodules at the left lung base. Despite administration of empiric antimicrobials for opportunistic bacterial, mycobacterial, and fungal pathogens commonly responsible for pulmonary complications of advanced HIV disease, the patient became and remained febrile to 39°C. A left-sided pleural effusion developed; diagnostic thoracentesis showed it to be exudative but otherwise non-diagnostic. Bronchoalveolar lavage performed three weeks later only demonstrated alveolar histiocytes, some of which were hemosiderin laden, a few lymphocytes, and neutrophils. FITE, AFB, and methanamine silver stains again were negative.

Two weeks thereafter, a wedge resection of the left lung was performed through CT-guided minithoracotomy. Multiple tissue sections revealed nodular areas of acute and chronic inflammation with regions of necrosis and fibrosis. Numerous multinuclated giant cells were present, some of which contained both intracytoplasmic and intranuclear inclusions suggestive of measles virus giant cell

- 77 -

pneumonia. Special stains for bacteria, fungi, P. carinii, and acid fast organisms again gave negative results. Electron microscopic examination of sections of this lung biopsy revealed particles morphologically consistent with paramyxoviruses such as measles virus. Serum anti-measles IgM titers determined by a solid phase hemadsorbant assay were negative, as was a subsequent IgM capture immunoassay.

5

10

15

20

25

30

35

Two weeks later, Rhesus monkey kidney (RMK) tissue culture cells inoculated with the patient's lung biopsy material revealed cytopathic changes characteristic of measles virus infection.

Confirmation was obtained using an immunofluorescence assay with monoclonal antibodies directed to measles virus. Based upon this diagnosis, oral ribavirin 1000mg B.I.D. was given for 14 days. Unfortunately, the patient progressively deteriorated, eventually dying two months later.

In order to ascertain the nature of the measles virus present in the patient, reverse transcription and PCR amplification of virus obtained from infected tissues were performed, followed by sequence analysis. The measles virus isolated from Rhesus monkey kidney cells inoculated with tissue from this patient's lung biopsy was propagated by two serial passages in the continuous Vero (monkey kidney) tissue culture cell line. Total infected cell RNA was extracted at the second Vero cell passage using TRIzol reagent (Life Technologies, Grand Island, NY) according to the manufacturer's protocol. Total RNA was similarly extracted from the patient's lung biopsy material. The measles virus vaccine strain (Moraten) currently used in the United States as a component of the trivalent MMR vaccines, was obtained in its univalent form (Attenuvax™, Merck, Sharpe, & Dohme).

- 78 -

This virus was passaged once in Vero cells and total vaccine infected cellular RNA then was extracted as described above.

5

10

15

20

25

30

Each of these RNA preparations was reverse transcribed (RT) to cDNA using random hexameric primers and Maloney murine leukemia virus reverse transcriptase (Perkin-Elmer/Cetus RT-PCR kit reagents, Perkin-Elmer-Cetus, Branchburg, NJ). The cDNA then was amplified by PCR using measles virus-specific oligodeoxynucleotide primer pairs whose design was based on the Edmonston measles virus sequence described above. These PCR products comprised a set of overlapping DNA fragments spanning the entire 15,894 nucleotide long measles genome. A consensus genomic sequence was established by direct analysis of each PCR product, without cloning, using the dideoxy terminator cycle-sequencing method established by the manufacturer (ABI PRISM 377 sequencer and ABI PRISM DNA sequencing kit; Perkin-Elmer/Cetus, Foster City, CA). Both strands of the PCR-amplified DNA products were analyzed to eliminate possible sequencing ambiguities.

The nucleotide sequences of selected regions of the measles virus genomes present in the patient's viral isolate, as well as in the diseased lung tissue, were compared with that of the Moraten vaccine virus, as well as with the nucleotide sequences of other measles virus wild-type and vaccine strains. This sequence analysis revealed identity to the Moraten vaccine strain rather than demonstrating relatedness to past or currently circulating wild-type viruses or other measles vaccine strains.

- 79 -

# Example 5 ELISA to Detect RSV

of RSV. Peptides are designed and selected based on homologies to the RSV sequences described herein to be specific for all subgroup B strains, or for individual wild-type, vaccine or revertant RSV subgroup B strains described herein. These peptides are then coupled to KLH and used to immunize rabbits for the production of monospecific polyclonal antibody. A selection of these polyclonal antibodies, or a combination of polyclonal and monoclonal antibodies is then used in a "capture ELISA" to detect the presence of an RSV antigen.

- 80 -

## Bibliography

- 1. Kapikian, A.Z., et al., Am. J.
- Epidemol., 89, 405-421 (1969).
- 2. Chin, J., et al., Am. J. Epidemol., 89, 449-463 (1969).
  - 3. Fulginiti, V.A., et al., Am. J.
- Epidemol., 89, 435-448 (1969).
- 4. Prince, G.A., et al., <u>J. Virology</u>, <u>57</u>, 721-728 (1986).
- 5. Kim, H.W., et al., <u>Pediatrics</u>, <u>52</u>, 56-63 (1973).
- 6. Hodes, D.S., et al., <u>Proc. Soc. Exp.</u> <u>Biol. Med.</u>, <u>145</u>, 1158-1164 (1974).
- 7. Belshe, R.B., and Hissom, F.K., <u>J. Med.</u> <u>Virol.</u>, <u>10</u>, 235-242 (1982).
- 8. Black, F.L., et al., Am. J. Epidemiol., 124, 442-452 (1986).
- 9. Lennon, J.L., and Black, F.L., <u>J.</u>
  <u>Pediatrics</u>, <u>108</u>, 671-676 (1986).
- 10. Pabst, H.F., et al., <u>Pediatr. Infect.</u> <u>Dis. J.</u>, <u>11</u>, 525-529 (1992).
- 11. Centers for Disease Control,  $\underline{MMWR}$ ,  $\underline{40}$ , 369-372 (1991).
- 12. Centers for Disease Control, MMWR, 41:S6, 1-12 (1992).
- 13. King, G.E., et al., <u>Pediatr. Infect.</u> <u>Dis. J.</u>, <u>10</u>, 883-887 (1991).
- 14. Rota, J.S., et al., <u>Virology</u>, <u>188</u>, 135-142 (1992).
- 15. Rota, J.S., et al., <u>Virus Res.</u>, <u>31</u>, 317-330 (1994).
- 16. Lamb, R.A., and Kolakosky, D., pages 1177-1204 of Volume 1, <u>Fields Virology</u>, B.N. Fields, et al., Eds. (3rd ed., Raven Press, 1996).

PCT/US97/16718

- 81 -

- 17. Sidhu, M.S., et al., <u>Virology</u>, <u>193</u>, 50-65 (1993).
- 18. Garcin, D., et al., <u>EMBO J.</u>, <u>14</u>, 6087-6094 (1995).
- 19. Radecke, F., et al., <u>EMBO J.</u>, <u>14</u>, 5773-5783 (1995).
- 20. Collins, P.L., et al., <u>Proc. Natl. Acad.</u> <u>Sci., USA</u>, <u>92</u>, 11563-11567 (1995).
- 21. Published European Patent Application No. 702,085.
- 22. Published International Application No. WO 96/10400.
- 23. Baron, M.D., and Barrett, T., <u>J.</u> <u>Virology</u>, <u>71</u>, 1265-1271 (1997).
- 24. Published International Application No. WO 97/06270.
- 25. U.S. Provisional Patent Application No. 60/047575.
- 26. Published International Application No. WO 97/12032.
- 27. Kato, A., et al., <u>Genes to Cells</u>, <u>1</u>, 569-579 (1996).
- 28. Sidhu, M.S., et al., <u>Virology</u>, <u>208</u>, 800-807 (1995).
- 29. Shaffer, M.F., et al., <u>J. Immunol.</u>, <u>41</u>, 241-256 (1941).
- 30. Enders, J.F., et al., N. Engl. J. Med., 263, 153-159 (1960).
- 31. Enders, J.F., and Peebles, M.E., <u>Proc.</u>
  <u>Soc. Exp. Biol. Med.</u>, <u>86</u>, 227-286 (1954).
- 32. Schwarz, A.J.F., Am. J. Dis. Child., 103, 216-219 (1962).
- 33. Griffin, D.E., and Bellini, W.J., pages 1267-1312 of Volume 1, <u>Fields Virology</u>, B.N. Fields, et al., Eds. (3rd ed., Raven Press, 1996).

PCT/US97/16718

WO 98/13501

- 82 -

- 34. Birrer, M.J., et al., Virology, 108, 381-390 (1981).
- 35. Birrer, M.J., et al., Nature, 293, 67-69 (1981).
- Norby, E., et al., pages 481-507, in The 36. Paramyxoviruses, D. Kingsbury, Ed. (Plenum Press, 1991).
- 37. Peebles, M.E., pages 427-456, in The Paramyxoviruses, D. Kingsbury, Ed. (Plenum Press, 1991).
- 38. Egelman, E.H., et al., J. Virol., 63, 2233-2243 (1989).
- 39. Udem, S.A., et al., J. Virol. Methods, 8, 123-136 (1984).
- 40. Udem, S.A., and Cook, K.A., J. Virol., 49, 57-65 (1984).
- 41. Moyer, S.A., and Horikami, S.M., pages 249-274, in The Paramyxoviruses, D. Kingsbury, Ed. (Plenum Press, 1991).
- Blumberg, B., et al., pages 235-247, in The Paramyxoviruses, D. Kingsbury, Ed. (Plenum Press, 1991).
- 43. Berrett, T., et al., pages 83-102, in The Paramyxoviruses, D. Kingsbury, Ed. (Plenum Press, 1991).
- 44. Tordo, N., et al., Sem. in Virology, 3, 341-357 (1992).
- 45. Cattaneo, R., et al., EMBO J., 6, 681-688 (1987).
- 46. Crowley, J.C., et al., Virology, 164, 498-506 (1988).
- 47. Banerjee, A.K., and Barik, S., et al., <u>Virology</u>, <u>188</u>, 417-428 (1992).
- 48. Castaneda, S.J., and Wong, T.C., J. Virol., 63, 2977-2986 (1989).

- 49. Chan, J., et al., pages 221-231, in Genetics and Pathogenicity of Negative Stranded Viruses, B.W.J. Mahy and D. Kolakofsky, Eds. (Elsevier Biomedical Press, 1989).
- 50. Blumberg, B., et al., <u>Cell</u>, <u>23</u>, 837-845 (1981).
- 51. Blumberg, B., et al., <u>Cell</u>, <u>32</u>, 559-567 (1983).
- 52. Kolakofsky, D., and Blumberg, B.M., pages 203-213, in <u>Virus Persistence</u>, B.M.J. Mahy, et al., Eds. (Cambridge University Press, 1982).
- 53. Castaneda, S.J., and Wong, T.C., <u>J.</u> <u>Virol.</u>, <u>64</u>, 222-230 (1990).
- 54. Curran, J.A., and Kolakofsky, D., Virology, 182, 168-176 (1991).
- 55. Sidhu, M.S., et al., <u>Virology</u>, <u>193</u>, 66-72 (1993).
- 56. Sidhu, M.S., et al., <u>Virology</u>, <u>202</u>, 631-641 (1994).
- 57. Collins, P.L., et al., pages 1205-1241 of Volume 1, <u>Fields Virology</u>, B.N. Fields, et al., Eds. (3rd ed., Raven Press, 1996).
- 58. Crookshanks, F.K., and Belshe, R.B., <u>J.</u>
  <u>Med. Virol.</u>, <u>13</u>, 243-249 (1984).
- 59. Crookshanks-Newman, F.K., and Belshe, R.B., <u>J. Med. Virol.</u>, <u>18</u>, 131-137 (1986).
- 60. Hall, S.L., et al., <u>Virus Res.</u>, <u>22</u>, 173-184 (1992).
- 61. Karron, R.A., et al., <u>J. Inf. Dis.</u>, <u>172</u>, 1445-1450 (1995).
- 62. Anderson, L.J., et al., <u>J. Infect. Dis.</u>, <u>151</u>, 626-633 (1985).
- 63. Collins, P.L., pages 103-162 of <u>The Paramyxoviruses</u>, D.W. Kingsbury, Ed. (Plenum Press, NY and London, 1991).

- 64. Sullender, W.M., <u>J. Virology</u>, <u>65</u>, 5425-5434 (1991).
- 65. Lerch, R.A., et al., <u>J. Virology</u>, <u>64</u>, 5559-5569 (1990).
- 66. Mallipeddi, S.K., and Samal, S.K., <u>J.</u> <u>Gen Virol.</u>, <u>74</u>, 2787-2791 (1993).
- 67. Johnson, P.R., et al., <u>J. Virology</u>, <u>61</u>, 3163-3166 (1987).
- 68. Stott, E.J., et al., <u>J. Virology</u>, <u>61</u>, 3855-3861 (1987).
- 69. Henderson, F.W., et al., N. Engl. J. Med., 300, 530-534 (1979).
- 70. Hall, S.L., et al., <u>J. Infect. Dis.</u>, <u>163</u>, 693-698 (1991).
- 71. Mufson, M.A., et al., <u>J. Gen. Virol.</u>, <u>66</u>, 2111-2124 (1985).
- 72. Glezen, W.P., et al., <u>Am. J. Dis.</u> Child., <u>140</u>, 543-546 (1986).
- 73. Hemming, V.G., et al., <u>Clin. Microbiol.</u>
  <u>Res.</u>, <u>8</u>, 22-33 (1995).
- 74. Collins, P. L. et. al., pages 1313-1351 of volume 1, <u>Fields Virology</u>, B. N. Fields, et al., Eds. (3rd ed., Raven Press, 1996).
- 75. Ling, R., and Pringle, C.R., <u>J. Gen.</u> <u>Virol.</u>, <u>70</u>, 1427-1440 (1989).
- 76. Yu, Q., et al., <u>J. Virology</u>, <u>69</u>, 2412-2419 (1995).
- 77. McIntosh, K., and Chanock, R.M., pages 1045-1072 of <u>Virology</u>, B.N. Fields, et al., Eds. (2nd ed., Raven Press, 1990).
- 78. Heminway, B.R., et al., page 167 of Abstracts of the IX International Congress of Virology, P17-2, (1993).
- 79. Mink, M.A., et al., <u>Virology</u>, <u>185</u>, 615-624 (1991).

- 80. Dickens, L.E., et al., <u>J. Virology.</u>, <u>52</u>, 364-369 (1990).
- 81. Wagner, R.R., and Rose, J.K., pages 1121-1135 of volume 1, <u>Fields Virology</u>, B.N. Fields, et al., Eds. (3rd ed., Raven Press, 1996).
- 82. Barik, S., <u>J. Gen. Virol.</u>, <u>74</u>, 485-490 (1993).
- 83. Collins, P.L., et al., pages 259-264 of Vaccines 93: modern approaches to new vaccines including prevention of AIDS, F. Brown et al., Eds. (Cold Spring Harbor Laboratory Press, NY, 1993).
- 84. Kuo, L., et al., <u>J. Virology.</u>, <u>70</u>, 6892-6901 (1996).
- 85. Huang, Y.T., and Wertz, G.W., <u>J.</u> <u>Viroloogy</u>, <u>43</u>, 150-157 (1982).
- 86. Sambrook, J., et al., Molecular Cloning:

  A Laboratory Manual, 2nd ed., Cold Spring Harbor

  Laboratory Press, Cold Spring Harbor, N.Y. (1989).
- 87. Ray, R., et al., <u>J. Virol.</u>, <u>69</u>, 1959-1963 (1995).
- 88. Ray, R., et al., <u>J. Virol.</u>, <u>70</u>, 580-584 (1996).
- 89. Stokes, A., et al., <u>Virus Research</u>, <u>30</u>, 43-52 (1993).
  - 90. U.S. Patent Application No. 08/059,444.

- 86 -

#### SEQUENCE LISTING

#### (1) GENERAL INFORMATION:

(i) APPLICANT: Udem, Stephen A.
Sidhu, Mohinderjit S.
Tatem, Joanne M.
Murphy, Brian R.
Randolph, Valerie B.

- (ii) TITLE OF INVENTION: 3' Genomic Promoter Region and Polymerase Gene Mutations Responsible for Attenuation in Viruses of the Order Designated Mononegavirales
- (iii) NUMBER OF SEQUENCES: 79
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: American Home Products Corporation
  - (B) STREET: One Campus Drive
  - (C) CITY: Parsippany
  - (D) STATE: New Jersey
  - (E) COUNTRY: United States
  - (F) ZIP: 07054
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Gordon, Alan M.
  - (B) REGISTRATION NUMBER: 30,637
  - (C) REFERENCE/DOCKET NUMBER: 33,294 PCT
  - (1x) TELECOMMUNICATION INFORMATION:
    - (A) TELEPHONE: 973/683-2157
    - (B) TELEFAX: 973/683-4117
- (2) INFORMATION FOR SEQ ID NO:1:
  - (1) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 15894 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single

- 87 -

## (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: RNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACCAAACAAA	GTTGGGTAAG	GATAGATCAA	TCAATGATCA	TATTCTAGTG	CACTTAGGAT	60
TCAAGATCCT	ATTATCAGGG	ACAAGAGCAG	GATTAGGGAT	ATCCGAGATG	GCCACACTTT	120
TAAGGAGCTT	AGCATTGTTC	AAAAGAAACA	AGGACAAACC	ACCCATTACA	TCAGGATCCG	180
GTGGAGCCAT	CAGAGGAATC	AAACACATTA	TTATAGTACC	AATCCCTGGA	GATTCCTCAA	240
TTACCACTCG	ATCCAGACTT	CTGGACCGGT	TGGTCAGGTT	AATTGGAAAC	CCGGATGTGA	300
GCGGGCCCAA	ACTAACAGGG	GCACTAATAG	GTATATTATC	CTTATTTGTG	GAGTCTCCAG	360
GTCAATTGAT	TCAGAGGATC	ACCGATGACC	CTGACGTTAG	CATAAGGCTG	TTAGAGGTTG	420
TCCAGAGTGA	CCAGTCACAA	TCTGGCCTTA	CCTTCGCATC	AAGAGGTACC	AACATGGAGG	480
ATGAGGCGGA	CCAATACTTT	TCACATGATG	ATCCAATTAG	TAGTGATCAA	TCCAGGTTCG	540
GATGGTTCGA	GAACAAGGAA	ATCTCAGATA	TTGAAGTGCA	AGACCCTGAG	GGATTCAACA	600
TGATTCTGGG	TACCATCCTA	GCCCAAATTT	GGGTCTTGCT	CGCAAAGGCG	GTTACGGCCC	660
CAGACACGGC	AGCTGATTCG	GAGCTAAGAA	GGTGGATAAA	GTACACCCAA	CAAAGAAGGG	720
TAGTTGGTGA	ATTTAGATTG	GAGAGAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC	CTTACGCCGA	TTCATGGTCG	CTCTAATCCT	GGATATCAAG	AGAACACCCG	840
GAAACAAACC	CAGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATC	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCCTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGGT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAC	CTTTACCAGC	1020
AAATGGGGGA	AACTGCACCC	TACATGGTAA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATC	ATACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG	AGGTTTGAAC	TTTGGCCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT	GGTAAGGAGG	TCAGCTGGAA	AGGTCAGTTC	CACATTGGCA	TCTGAACTCG	1260
GTATCACTGC	CGAGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCATACT	ACTGAGGACA	1320

- 88 -

AGATCAGTAG	AGCGGTTGGA	CCCAGACAAG	CCCAAGTATC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA	GCTACCGAGA	TTGGGGGGCA	AGGAAGATAG	GAGGGTCAAA	CAGAGTCGAG	1440
GAGAAGCCAG	GGAGAGCTAC	AGAGAAACCG	GGCCCAGCAG	AGCAAGTGAT	GCGAGAGCTG	1500
CCCATCTTCC	AACCGGCACA	CCCCTAGACA	TTGACACTGC	ATCGGAGTCC	AGCCAAGATC	1560
CGCAGGACAG	TCGAAGGTCA	GCTGACGCCC	TGCTTAGGCT	GCAAGCCATG	GCAGGAATCT	1620
CGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTATAGTGTA	CAATGACAGA	AATCTTCTAG	1680
ACTAGGTGCG	AGAGGCCGAG	GACCAGAACA	ACATCCGCCT	ACCCTCCATC	ATTGTTATAA	1740
AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCCATCA	ACCATCCACT	CCCACGATTG	1800
GAGCCGATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	ATCGAGGAAG	CTATGGCAGC	ATGGTCAGAA	1920
ATATCAGACA	ACCCAGGACA	GGAGCGAGCC	ACCTGCAGGG	AAGAGAAGGC	AGGCAGTTCG	1980
GGTCTCAGCA	AACCATGCCT	CTCAGCAATT	GGATCAACTG	AAGGCGGTGC	ACCTCGCATC	2040
CGCGGCCAGG	GACCTGGAGA	GAGCGATGAC	GACGCTGAAA	CTTTGGGAAT	CCCCCAAGA	2100
AATCTCCAGG	CATCAAGCAC	TGGGTTACAG	TGTTATTATG	TTTATGATCA	CAGCGGTGAA	2160
GCGGTTAAGG	GAATCCAAGA	TGCTGACTCT	ATCATGGTTC	AATCAGGCCT	TGATGGTGAT	2220
AGCACCCTCT	CAGGAGGAGA	CAATGAATCT	GAAAACAGCG	ATGTGGATAT	TGGCGAACCT	2280
GATACCGAGG	GATATGCTAT	CACTGACCGG	GGATCTGCTC	CCATCTCTAT	GGGGTTCAGG	2340
GCTTCTGATG	TTGAAACTGC	AGAAGGAGGG	GAGATCCACG	AGCTCCTGAG	ACTCCAATCC	2400
AGAGGCAACA	ACTTTCCGAA	GCTTGGGAAA	ACTCTCAATG	TTCCTCCGCC	CCCGGACCCC	2460
GGTAGGGCCA	GCACTTCCGA	GACACCCATT	AAAAAGGGCA	CAGACGCGAG	ATTAGCCTCA	2520
TTTGGAACGG	AGATCGCGTC	TTTATTGACA	GGTGGTGCAA	CCCAATGTGC	TCGAAAGTCA	2580
CCCTCGGAAC	CATCAGGGCC	AGGTGCACCT	GCGGGGAATG	TCCCCGAGTG	TGTGAGCAAT	2640
GCCGCACTGA	TACAGGAGTG	GACACCCGAA	TCTGGTACCA	CAATCTCCCC	GAGATCCCAG	2700
DAADTAATAA	AAGGGGGAGA	CTATTATGAT	GATGAGCTGT	TCTCTGATGT	CCAAGATATT	2760
AAAACAGCCT	TGGCCAAAAT	ACACGAGGAT	AATCAGAAGA	TAATCTCCAA	GCTAGAATCA	2820
CTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAGAAGC	AGATCAACAG	GCAAAATATC	2880

- 89 -

AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TCGCCATTC	TGGACTTGGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTO	GAAATCAATC	CCGACTTGA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AACCCGTTG	CAGCCGACAA	3060
CTCCAAGGAA	TGACAAATGG	ACGGACCAGI	TCCAGAGGAC	AGCTGCTGA	GGAATTTCAG	3120
CTAAAGCCGA	TCGGGAAAAA	GATGAGCTCA	GCCGTCGGGT	TTGTTCCTG	CACCGGCCCT	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGA	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ATGATCTTGC	CAAGTTCCAC	3300
CAGATGCTGA	TGAAGATAAT	AATGAAGTAG	CTACAGCTCA	ACTTACCTGC	CAACCCCATG	3360
CCAGTCGACC	CAACTAGTAC	AACCTAAATC	CATTATAAAA	AACTTAGGAG	CAAAGTGATT	3420
GCCTCCCAAG	TTCCACAATG	ACAGAGATCT	ACGACTTCGA	CAAGTCGGCA	TGGGACATCA	3480
AAGGGTCGAT	CGCTCCGATA	CAACCCACCA	CCTACAGTGA	TGGCAGGCTG	GTGCCCCAGG	3540
TCAGAGTCAT	AGATCCTGGT	CTAGGCGACA	GGAAGGATGA	ATGCTTTATG	TACATGTTTC	3600
TGCTGGGGGT	TGTTGAGGGC	AGCGATCCCC	TAGGGCCTCC	AATCGGGCGA	GCATTTGGGT	3660
CCCTGCCCTT	AGGTGTTGGC	AGATCCACAG	CAAAGCCCGA	AGAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTGA	CATAGTTGTT	AGACGTACAG	CAGGGCTCAA	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCCC	ACTAACTCTC	CTCACACCTT	GGAGAAAGGT	CCTAACAACA	GGGAGTGTCT	3840
TCAACGCAAA	CCAAGTGTGC	AATGCGGTTA	ATCTGATACC	GCTCGATACC	CCGCAGAGGT	3900
TCCGTGTTGT	TTATATGAGC	ATCACCCGTC	TTTCGGATAA	CGGGTATTAC	ACCGTTCCTA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	GGCGATAGGC	CCTGGGAAGA	TCATCGACAA	TACAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	GAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AGATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCTTATGT	TACCCGCTGA	TGGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380
AAGAATTCCG	CATTTACGAC	GACGTGATCA	TAAATGATGA	CCAAGGACTA	TTCAAAGTTC	4440

- 90 -

TGTAGACCGT	AGTGCCCAGC	AATGCCCGAA	AACGACCCCC	CTCACAATGA	CAGCCAGAAG	4500
GCCCGGACAA	AAAAGCCCCC	TCCGAAAGAC	TCCACGGACC	AAGCGAGAGG	CCAGCCAGCA	4560
GCCGACGGCA	AGCGCGAACA	CCAGGCGGCC	CCAGCACAGA	ACAGCCCTGA	CACAAGGCCA	4620
CCACCAGCCA	CCCCAATCTG	CATCCTCCTC	GTGGGACCCC	CGAGGACCAA	CCCCCAAGGC	4680
TGCCCCCGAT	CCAAACCACC	AACCGCATCC	CCACCACCCC	CGGGAAAGAA	ACCCCCAGCA	4740
ATTGGAAGGC	CCCTCCCCCT	CTTCCTCAAC	ACAAGAACTC	CACAACCGAA	CCGCACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGATCCTCTC	TCCCCGGCAA	4860
ACTAAACAAA	ACTTAGGGCC	AAGGAACATA	CACACCCAAC	AGAACCCAGA	CCCCGGCCCA	4920
cgccccccc	CCCCCAACCC	CCGACAACCA	GAGGGAGCCC	CCAACCAATC	CCGCCGGTTC	4980
CCCCGGTGCC	CACAGGCAGG	GACACCAACC	CCCGAACAGA	CCCAGCACCC	AACCATCGAC	5040
AATCCAAGAC	GGGGGGCCC	CCCCAAAAAA	AGGCCCCCAG	GGGCCGACAG	CCAGCACCGC	5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CGGCAACCAA	ACCAGAACCC	AGACCACCCT	5160
GGGCCACCAG	CTCCCAGACT	CGGCCATCAC	CCCGCAGAAA	GGAAAGGCCA	CAACCCGCGC	5220
ACCCCAGCCC	CGATCCGGCG	GGGAGCCACC	CAACCCGAAC	CAGCACCCAA	GAGCGATCCC	5280
CGAAGGACCC	CCGAACCGCA	AAGGACATCA	GTATCCCACA	GCCTCTCCAA	GTCCCCCGGT	5340
CTCCTCCTTT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CACCCGACGA	CACTCAACTC	5400
CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAGACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGAAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTAG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGGTT	5760
CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCGG	GAGTAGTCCT	GGCAGGTGCG	5820
GCCCTAGGCG	TTGCCACAGC	TGCTCAGATA	ACAGCCGGCA	TTGCACTTCA	CCAGTCCATG	5880
CTGAACTCTC	AAGCCATCGA	CAATCTGAGA	GCGAGCCTGG	AAACTACTAA	TCAGGCAATT	5940
GAGGCAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000

- 91 -

ATCAATAATG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTCGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCCAGCTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	TGGAGGAGAC	6180
ATCAATAAGG	TGTTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
AGTATAGCCT	ATCCGACGCT	GTCCGAGATT	AAGGGGGTGA	TTGTCCACCG	GCTAGAGGGG	6360
GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420
CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGGG	6540
TCCACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGACGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGCTCCTCCC	ATATCATTGG	AGAGGTTGGA	CGTAGGGACA	6840
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	AATTGTTGGA	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTATCG	AGCACTAGCA	TAGTCTACAT	CCTGATTGCA	6960
GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACGGGA	7080
ACATCARART	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCACCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	AGTAGGATAG	TCATTAACAG	AGAACATCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTTG	TCATGTCTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560

- 92 -

2	AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
(	CATCTCTGAC	AAGATTAAAT	TCCTTAATCC	GGATAGGGAG	TACGACTTCA	GAGATCTCAC	7680
•	ITGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
(	GCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
(	CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
1	ATTCTCAAAC	ATGTCGCTGT	CCCTGTTAGA	CTTGTATTTA	AGTCGAGGTT	ACAATGTGTC	7920
1	ATCTATAGTC	ACTATGACAT	CCCAGGGAAT	GTATGGGGGA	ACTTACCTAG	TGGAAAAGCC	7980
•	FAATCTGAGC	AGCAAAAGGT	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040
1	AGGTGTTATC	AGAAATCCGG	GTTTGGGGGC	TCCGGTGTTC	CATATGACAA	ACTATCTTGA	8100
(	GCAACCAGTC	AGTAATGATC	TCAGCAACTG	TATGGTGGCT	TTGGGGGAGC	TCAAACTCGC	8160
1	AGCCCTTTGT	CACGGGGAAG	ATTCTATCAC	AATTCCCTAT	CAGGGATCAG	GGAAAGGTGT	8220
(	CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCTG	GAAATCCCCA	ACCGACATGC	AATCCTGGGT	8280
(	CCCTTATCA	ACGGATGATC	CAGTGATAGA	CAGGCTTTAC	CTCTCATCTC	ACAGAGGTGT	8340
•	PATCGCTGAC	AATCAAGCAA	AATGGGCTGT	CCCGACAACA	CGAACAGATG	ACAAGTTGCG	8400
1	AATGGAGACA	TGCTTCCAAC	AGGCGTGTAA	GGGTAAAATC	CAAGCACTCT	GCGAGAATCC	8460
(	CGAGTGGGCA	CCATTGAAGG	ATAACAGGAT	TCCTTCATAC	GGGGTCTTGT	CTGTTGATCT	8520
(	GAGTCTGACA	GTTGAGCTTA	AAATCAAAAT	TGCTTCGGGA	TTCGGGCCAT	TGATCACACA	8580
1	CGGTTCAGGG	ATGGACCTAT	ACAAATCCAA	CCACAACAAT	GTGTATTGGC	TGACTATCCC	8640
•	GCCAATGAAG	AACCTAGCCT	TAGGTGTAAT	CAACACATTG	GAGTGGATAC	CGAGATTCAA	8700
•	GGTTAGTCCC	AACCTCTTCA	CTGTCCCAAT	TAAGGAAGCA	GGCGAAGACT	GCCATGCCCC	8760
1	AACATACCTA	CCTGCGGAGG	TGGATGGTGA	TGTCAAACTC	AGTTCCAATC	TGGTGATTCT	8820
1	ACCTGGTCAA	GATCTCCAAT	ATGTTTTGGC	AACCTACGAT	ACTTCCAGGG	TTGAACATGC	8880
•	IGTGGTTTAT	TACGTTTACA	GCCCAGGCCG	CTCATTTTCT	TACTTTTATC	CTTTTAGGTT	8940
•	GCCTATAAAG	GGGGTCCCCA	TCGAATTACA	AGTGGAATGC	TTCACATGGG	ACCAAAAACT	9000
•	CTGGTGCCGT	CACTTCTGTG	TGCTTGCGGA	CTCAGAATCT	GGTGGACATA	TCACTCACTC	9060
•	TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACCCGG	GAAGATGGAA	CCAATCGCAG	9120

- 93 -

ATAGGGCTGC	: TAGTGAACCA	ATCACATGAT	GTCACCCAGA	CATCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATI	* AAGAAAAACG	TAGGGTCCAA	GTGGTTCCCC	GTTATGGACT	9240
CGCTATCTGT	CAACCAGATO	TTATACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
ATAAGATAGT	AGCCATCCTG	GAGTATGCTC	GAGTCCCTCA	CGCTTACAGO	CTGGAGGACC	9360
CTACACTGTG	TCAGAACATC	AAGCACCGCC	TAAAAAACGG	ATTTTCCAAC	CAAATGATTA	9420
TAAACAATGT	GGAAGTTGGG	AATGTCATCA	AGTCCAAGCT	TAGGAGTTAT	CCGGCCCACT	9480
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTTAA	CATAGAAGAC	AAAGAGTCAA	9540
CGAGGAAGAT	CCGTGAACTC	CTCAAAAAGG	GGAATTCGCT	GTACTCCAAA	GTCAGTGATA	9600
AGGTTTTCCA	ATGCTTAAGG	GACACTAACT	CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAGTT	ATTAACTIGG	GAGTTTACAT	GCACAGCTCC	CAGTGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TGACATTTGA	ACTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGAC	TATTGATGCT	AGGTATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AACTGATAGA	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCCATGC	10080
TGGAGCCTCT	TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTAA	CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
atgaaggtac	TTATCATGAG	TTAATTGAAG	CTCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGAC	AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGGC	TGCTGAAAAT	GTTAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGAT	GAAAGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAGG	CAGTTGGCCA	CCGCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAAGC	TTCAGGTGAA	GGGTTAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAÄATCTT	10560
					CTGACAATGT	10620
ACCTAAAGGA	CAAGGCACTT	GCTGCTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680

- 94 -

AGTTCCTGCG	TTACGACCCT	CCCAAGGGAA	CCGGGTCACG	GAGGCTTGTA	GATGTTTTCC	10740
TTAATGATTC	GAGCTTTGAC	CCATATGATG	TGATAATGTA	TGTTGTAAGT	GGAGCTTACC	10800
TCCATGACCC	TGAGTTCAAC	CTGTCTTACA	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTTTT	TGCTAAAATG	ACTTACAAAA	TGAGGGCATG	CCAAGTGATT	GCTGAAAATC	10920
TAATCTCAAA	CGGGATTGGC	ATATTTTA	AGGACÀATGG	GATGGCCAAG	GATGAGCACG	10980
ATTTGACTAA	GGCACTCCAC	ACTCTAGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGGGG	GGGGCCAGTC	TTAAAAACCT	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100
GGAACGTGAG	AGCAGCAAAA	GGGTTTATAG	GGTTCCCTCA	AGTAATTCGG	CAGGACCAAG	11160
ACACTGATCA	TCCGGAGAAT	ATGGAAGCTT	ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAGAA	GTACTGCCTT	AATTGGAGAT	ATGAGACCAT	CAGCTTGTTT	GCACAGAGGC	11280
TAAATGAGAT	TTACGGATTG	CCCTCATTTT	TCCAGTGGCT	GCATAAGAGG	CTTGAGACCT	11340
CTGTCCTGTA	TGTAAGTGAC	CCTCATTGCC	CCCCGACCT	TGACGCCCAT	ATCCCGTTAT	11400
ATAAAGTCCC	CAATGATCAA	ATCTTCATTA	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTATCTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGGAT	TGCTTCGTTA	GTGCAAGGGG	ACAATCAGAC	CATAGCCGTA	ACAAAAAGGG	11580
TACCCAGCAC	ATGGCCCTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ATATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAA	ATTCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACAACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGGG	ATGAATTATC	12060
TGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCCTCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180
CACAACAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240

- 95 -

TTGTATGTGT CCAGAGCATC ACTAGACTCC TCAAGAACAT AACTGCAAGG TTTGTCCTGA	12300
TCCATAGTCC AAACCCAATG TTAAAAGGAT TATTCCATGA TGACAGTAAA GAAGAGGACG	12360
AGGGACTGGC GGCATTCCTC ATGGACAGGC ATATTATAGT ACCTAGGGCA GCTCATGAAA	12420
TCCTGGATCA TAGTGTCACA GGGGCAAGAG AGTCTATTGC AGGCATGCTG GATACCACAA	12480
AAGGCCTGAT TCGAGCCAGC ATGAGGAAGG GGGGGTTAAC CTCTCGAGTG ATAACCAGAT	12540
TGTCCAATTA TGACTATGAA CAATTCAGAG CAGGGATGGT GCTATTGACA GGAAGAAAGA	12600
GAAATGTCCT CATTGACAAA GAGTCATGTT CAGTGCAGCT GGCGAGAGCT CTAAGAAGCC	12660
ATATGTGGGC GAGGCTAGCT CGAGGACGGC CTATTTACGG CCTTGAGGTC CCTGATGTAC	12720
TAGAATCTAT GCGAGGCCAC CTTATTCGGC GTCATGAGAC ATGTGTCATC TGCGAGTGTG	12780
GATCAGTCAA CTACGGATGG TTTTTTGTCC CCTCGGGTTG CCAACTGGAT GATATTGACA	12840
AGGAAACATC ATCCTTGAGA GTCCCATATA TTGGTTCTAC CACTGATGAG AGAACAGACA	12900
TGAAGCTTGC CTTCGTAAGA GCCCCAAGTC GATCCTTGCG ATCTGCTGTT AGAATAGCAA	12960
CAGTGTACTC ATGGGCTTAC GGTGATGATG ATAGCTCTTG GAACGAAGCC TGGTTGTTGG	13020
CTAGGCAAAG GGCCAATGTG AGCCTGGAGG AGCTAAGGGT GATCACTCCC ATCTCAACTT	13080
CGACTAATTT AGCGCATAGG TTGAGGGATC GTAGCACTCA AGTGAAATAC TCAGGTACAT	13140
CCCTTGTCCG AGTGGCGAGG TATACCACAA TCTCCAACGA CAATCTCTCA TTTGTCATAT	13200
CAGATAAGAA GGTTGATACT AACTTTATAT ACCAACAAGG AATGCTTCTA GGGTTGGGTG	13260
TTTTAGAAAC ATTGTTTCGA CTCGAGAAAG ATACCGGATC ATCTAACACG GTATTACATC	13320
TTCACGTCGA AACAGATTGT TGCGTGATCC CGATGATAGA TCATCCCAGG ATACCCAGCT	13380
CCCGCAAGCT AGAGCTGAGG GCAGAGCTAT GTACCAACCC ATTGATATAT GATAATGCAC	13440
CTTTAATTGA CAGAGATGCA ACAAGGCTAT ACACCCAGAG CCATAGGAGG CACCTTGTGG	13500
AATTTGTTAC ATGGTCCACA CCCCAACTAT ATCACATTTT AGCTAAGTCC ACAGCACTAT	13560
CTATGATTGA CCTGGTAACA AAATTTGAGA AGGACCATAT GAATGAAATT TCAGCTCTCA	13620
TAGGGGATGA CGATATCAAT AGTTTCATAA CTGAGTTTCT GCTCATAGAG CCAAGATTAT	13680
TCACTATCTA CTTGGGCCAG TGTGCGGCCA TCAATTGGGC ATTTGATGTA CATTATCATA	13740
GACCATCAGG GAAATATCAG ATGGGTGAGC TGTTGTCATC GTTCCTTTCT AGAATGAGCA	13800

- 96 -

AAGGAGTGTT	TAAGGTGCTT	GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
GGCATTGTGG	TATTATAGAG	CCTATCCATG	GTCCTTCACT	TGATGCTCAA	AACTTGCACA	13920
CAACTGTGTG	CAACATGGTT	TACACATGCT	ATATGACCTA	CCTCGACCTG	TTGTTGAATG	13980
AAGAGTTAGA	AGAGTTCACA	TTTCTCTTGT	GTGAAAGCGA	CGAGGATGTA	GTACCGGACA	14040
GATTCGACAA	CATCCAGGCA	AAACACTTAT	GTGTTCTGGC	AGATTTGTAC	TGTCAACCAG	14100
GGACCTGCCC	ACCAATTCGA	GGTCTAAGAC	CGGTAGAGAA	ATGTGCAGTT	CTAACCGACC	14160
ATATCAAGGC	AGAGGCTAGG	TTATCTCCAG	CAGGATCTTC	GTGGAACATA	AATCCAATTA	14220
TTGTAGACCA	TTACTCATGC	TCTCTGACTT	ATCTCCGGCG	AGGATCGATC	AAACAGATAA	14280
GATTGAGAGT	TGATCCAGGA	TTCATTTTCG	ACGCCCTCGC	TGAGGTAAAT	GTCAGTCAGC	14340
CAAAGATCGG	CAGCAACAAC	ATCTCAAATA	TGAGCATCAA	GGATTTCAGA	CCCCCACACG	14400
ATGATGTTGC	AAAATTGCTC	AAAGATATCA	ACACAAGCAA	GCACAATCTT	CCCATTTCAG	14460
GGGGCAATCT	CGCCAATTAT	GAAATCCATG	CTTTCCGCAG	AATCGGGTTG	AACTCATCTG	14520
CTTGCTACAA	AGCTGTTGAG	ATATCAACAT	TAATTAGGAG	ATGCCTTGAG	CCAGGGGAAG	14580
ACGGCTTGTT	CTTGGGTGAG	GGATCGGGTT	CTATGTTGAT	CACTTATAAG	GAGATACTTA	14640
AACTAAACAA	GTGCTTCTAT	AATAGTGGGG	TTTCCGCCAA	TTCTAGATCT	GGTCAAAGGG	14700
AATTAGCACC	CTATCCCTCC	GAAGTTGGCC	TTGTCGAACA	CAGAATGGGA	GTAGGTAATA	14760
TTGTCAAAGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	GTAGATTGCT	14820
TCAATTTCAT	AGTTAGTAAT	ATCCCTACCT	CTAGTGTGGG	GTTTATCCAT	TCAGATATAG	14880
AGACCTTGCC	TAACAAAGAT	ACTATAGAGA	AGCTAGAGGA	ATTGGCAGCC	ATCTTATCGA	14940
TGGCTCTGCT	CCTGGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
GGGATTTTGT	TCAGGGATTT	ATAAGTTATG	TAGGGTCTCA	TTATAGAGAA	GTGAACCTTG	15060
TATACCCTAG	ATACAGCAAC	TTCATATCTA	CTGAATCTTA	TTTGGTTATG	ACAGATCTCA	15120
AGGCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	GATAATTGAA	TCATCTGTGA	15180
GGACTTCACC	TGGACTTATA	GGTCACATCC	TATCCATTAA	GCAACTAAGC	TGCATACAAG	15240
CAATTGTGGG	AGACGCAGTT	AGTAGAGGTG	ATATCAATCC	TACTCTGAAA	AAACTTACAC	15300
CTATAGAGCA	GGTGCTGATC	AATTGCGGGT	TGGCAATTAA	CGGACCTAAG	CTGTGCAAAG	15360

- 97 -

AATTGATCCA CCATGATGTT GCCTCAGGGC AAGATGGATT GCTTAATTCT ATACTCATCC 15420 TCTACAGGGA GTTGGCAAGA TTCAAAGACA ACCAAAGAAG TCAACAAGGG ATGTTCCACG 15480 CTTACCCCGT ATTGGTAAGT AGCAGGCAAC GAGAACTTAT ATCTAGGATC ACCCGCAAAT 15540 TTTGGGGGCA CATTCTTCTT TACTCCGGGA ACAGAAAGTT GATAAATAAG TTTATCCAGA 15600 ATCTCAAGTC CGGCTATCTG ATACTAGACT TACACCAGAA TATCTTCGTT AAGAATCTAT 15660 CCAAGTCAGA GAAACAGATT ATTATGACGG GGGGTTTGAA ACGTGAGTGG GTTTTTAAGG 15720 TAACAGTCAA GGAGACCAAA GAATGGTATA AGTTAGTCGG ATACAGTGCC CTGATTAAGG 15780 ACTAATTGGT TGAACTCCGG AACCCTAATC CTGCCCTAGG TGGTTAGGCA TTATTTGCAA 15840 TATATTAAAG AAAACTTTGA AAATACGAAG TTTCTATTCC CAGCTTTGTC TGGT 15894

#### (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2183 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15

Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala 20 25 30

Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45

Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 55 60

Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80

Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn 85 90 95

- 98 -

Ile	Glu	Авр	Lys 100	Glu	Ser	Thr	Arg	Lys 105	Ile	Arg	Glu	Leu	Leu 110	Lys	Lys
Gly	Asn	Ser 115	Leu	Tyr	Ser	Lys ,	Val 120	Ser	λap	Lys	Val	Phe 125	Gln	Сув	Leu
Arg	Asp 130	Thr	Asn	Ser	Arg	Leu 135	Gly	Leu	Gly	Ser	Glu 140	Leu	Arg	Glu	Asp
11e 145	Lys	Glu	Lys	Val	11e 150		Leu	Gly	Val	Tyr 155	Met	Hie	Ser	Ser	Gln 160
Trp	Phe	Glu	Pro	Phe 165	Leu	Phe	Trp	Phe	Thr 170	Val	Lys	Thr	Glu	Met 175	Arg
Ser	Val	Ile	Lys 180	Ser	Gln	Thr	His	Thr 185	Сув	His	Arg	Arg	Arg 190	His	Thr
Pro	Val	Phe 195	Phe	Thr	Gly	Ser	Ser 200	Val	Glu	Leu	Leu	Ile 205	Ser	Arg	Asp
Leu	Val 210	Ala	Ile	Ile	Ser	Lys 215	Glu	Ser	Gln	His	Val 220	Tyr	Tyr	Leu	Thr
Phe 225	Glu	Leu	Val	Leu	Met 230	Tyr	Сув	Ąsp	Val	11e 235	Glu	Gly	Arg	Leu	Met 240
				245	Thr		_		250	•				255	
Arg	Val	Arg	Tyr 260	Met	Trp	Lys	Leu	11e 265	Asp	Gly	Phe	Phe	Pro 270	Ala	Leu
Gly	Asn	Pro 275	Thr	Tyr	Gln	Ile	Val 280	Ala	Met	Leu	Glu	Pro 285	Leu	Ser	Leu
Ala	Tyr 290	Leu	Gln	Leu	Arg	<b>Авр</b> 295	Ile	Thr	Val	Glu	Leu 300	Arg	Gly	Ala	Phe
305					Thr 310					315		_			320
Phe	Ser	Авр	Glu	Gly 325	Thr	Tyr	His	Glu	130	Ile	Glu	Ala	Leu	Авр 335	Tyr
Ile	Phe	Ile	Thr 340	Asp	Asp	Ile	His	Leu 345	Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
Phe	Arg	Ser 355	Phe	Gly	His	Pro	Arg 360	Leu	Glu	Ala	Val	Thr 365	Ala	λla	Glu
Asn	Val	Arg	Lys	Tyr	Met	Asn	Gln	Pro	Lys	Val	Ile	Val	Tyr	Glu	Thr

- 99 -

1	376	)				379	5				38	0			
<b>Let</b> 385	ı Met	L Ly	s Gl	y Hi	390		Phe	Су	s Gl	y Ile 39		e Ile	a Ası	ı Gl	Ty 40
Arg	, As <u>r</u>	Àr	y Hi	405	y Gly S	Ser	Trp	Pro	410		ı Thi	. Le	ı Pro	415	
Ala	Ala	A A B I	420	r Ile	Arg	) Asn	Ala	425		a Ser	Gl <sub>3</sub>	/ Glu	1 Gly 430		ı Th
His	Glu	435	ı Cya	Va]	qa <i>A</i>	Asn	Trp 440		Se:	Phe	Ala	Gly 445		. Lys	Ph
Gly	Сув 450	Phe	Met	: Pro	Leu	Ser 455		Asp	Ser	Asp	460		Met	Туг	Le
Lys 465	Asp	Lys	Ala	Leu	470		Leu	Gln	Arg	475		Авр	Ser	Val	131 480
Pro	Lys	Glu	Phe	Leu 485	Arg	Tyr	Asp	Pro	490		Gly	Thr	Gly	Ser 495	_
			500		Phe			505					510		
Val	Ile	Met 515	Tyr	Val	Val	Ser	Gly 520	Ala	Tyr	Leu	His	Авр 525	Pro	Glu	Phe
	530				Leu	535					540				
545					Thr 550					555					560
				565	Asn				570					575	
			580		His			585					590		
		595			Lys		600					605			
Val	Leu 610	Lys	Thr	Tyr	Ser	Arg 615	Ser	Pro	Val	His	Thr 620	Ser	Thr	Arg	Asn
525					Gly 630					635					640
Asp	Gln	Asp	Thr	Asp 645	His	Pro	Glu	Asn	Met 650	Glu	Ala	Tyr	Glu	Thr 655	Val

- 100 -

Ser	Ala	Phe	Ile 660	Thr	Thr	Asp	Leu	Lув 665	Lys	Tyr	Сув	Leu	<b>Asn</b> 670	Trp	Arg
Tyr	Glu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685	Ile	Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	100 700	Glu	Thr	Ser	Val
705					Pro 710			٠		715					720
Pro	Leu	Tyr	Lys	Val 725	Pro	Asn	Авр	Gln	11e 730	Phe	Ile	Lys	Tyr	Pro 735	Met
			740		Tyr			745					750		
		755			Ala		760					765			
Leu	Val 770	Gln	Gly	Asp	Asn	Gln 775	Thr	Ile	Ala	Val	Thr 780	Lys	Arg	Val	Pro
785					Asn 790					795					800
				805	Ile				810					815	
			820		Glu			825					830		-
		835			Tyr		840					845			
	850				Val	855					860				
865					Asn 870					875					880
				885	Tyr				890					895	
			900		Ile			905					910		
Thr	Arg	<b>А</b> вр 915	Val	Val	Ile	Pro	Leu 920	Leu	Thr	Asn	Asn	Asp 925	Leu	Leu	Ile

- 101 -

Arg	Met 930	: Ala	Leu	Leu	Pro	Ala 935		Ile	Gly	gly	940		Туг	r Lev	a Asn
Met 945	Ser	Arg	Leu	Phe	Val 950	Arg	Asn	Ile	Gly	955		Val	. Thr	Ser	960
Ile	Ala	Asp	Leu	Lys 965	Arg	Met	Ile	Leu	Ala 970		Leu	Met	Pro	975	Glu
Thr	Leu	. His	Gln 980	Val	Met	Thr	Gln	Gln 985		Gly	Asp	Ser	Ser 990		Leu
Asp	Trp	Ala 995	Ser	Asp	Pro	Tyr	Ser 100		Asn	Leu	Val	Сув 100		Gln	Ser
Ile	Thr 101	Arg 0	Leu	Leu	ГÀÈ	Asn 101		Thr	Ala	Arg	Phe 102		Leu	Ile	His
Ser 102!	Pro 5	Asn	Pro	Met	Leu 103(		Gly	Leu	Phe	His 103		Asp	Ser	Lys	Glu 1040
Glu	Asp	Glu	Gly	Leu 1045		Ala	Phe	Leu	Met 105		Arg	His	Ile	11e 105	
Pro	Arg	Ala	Ala 1060		Glu	Ile	Leu	Asp 1069		Ser	Val	Thr	Gly 107		Arg
Glu	Ser	Ile 1075		Gly	Met	Leu	Asp 108(		Thr	Lys	Gly	Leu 108!		Arg	Ala
Ser	<b>Met</b> 1090	Arg O	Lys	Gly	Gly	Leu 1095		Ser	Arg	Val	Ile 1100		Arg	Leu	Ser
Asn 1105	Tyr	Asp	Tyr	Glu	Gln 1110		Arg	Ala	Gly	Met 1115		Leu	Leu	Thr	Gly 1120
Arg	Lys	Arg	Asn	Val 1125	Leu	Ile	Asp	Lys	Glu 1130		Сув	Ser	Val	Gln 1135	
Ala	Arg	Ala	Leu 1140	Arg	Ser	His		Trp 1145		Arg	Leu	Ala	Arg 115(		Arg
Pro	Ile	Tyr 1155	Gly	Leu	Glu	Val	Pro 1160	Asp	Val	Leu	Glu	Ser 1165		Arg	Gly
lis	Leu 1170	Ile	Arg .	Arg 1	His	Glu 1175	Thr	Сув	Val	Ile	Сув 1180		Сув	Gly	Ser
7al .	Asn	Tyr	Gly	Trp :	Phe :	Phe	Val	Pro		Gly 1195		Gln	Leu		Asp

Ile Asp Lys Glu Thr Ser Ser Leu Arg Val Pro Tyr Ile Gly Ser Thr

- 102 -

				1203	1				1210	,				1773	•
Thr	Asp	Glu	Arg 1220		Asp	Met	-	Leu 1225		Phe	Val		Ala 1230		Ser
Arg	Ser	Leu 1235	_	Ser	Ala	Val	Arg 1240		Ala	Thr	Val	Tyr 1245		Trp	Ala
Tyr	Gly 1250	-	Asp	Asp	Ser	Ser 1255		Asn	Glu	Ala	Trp 1260		Leu	Ala	Arg
Gln 1265	_	Ala	Asn	Val	Ser 1270	Leu	Glu	Glu	Leu	Arg 1275		Ile	Thr	Pro	Ile 1280
Ser	Thr	Ser	Thr	Asn 1285		Ala	His	Arg	Leu 1290	_	Asp	Arg	Ser	Thr 1295	
Val	Lys	Tyr	Ser 1300	_	Thr	Ser	Leu	Val 1305	_	Val	Ala	Arg	Tyr 1310		Thr
Ile	Ser	Asn 1315	_	Asn	Leu	Ser	Phe 1320		Ile	Ser	yab	Lув 1325	_	Val	Asp
Thr	Asn 1330		Ile	Tyr	Gln	Gln 1335	_	Met	Leu	Leu	Gly 1340		Gly	Val	Leu
Glu 134		Leu	Phe	Arg	Leu 1350	Glu )	Lys	Asp	Thr	Gly 1355		Ser	Asn	Thr	Val 1360
Leu	His	Leu	His	Val 136		Thr	Asp	Cys	Сув 1370		Ile	Pro	Met	Ile 1375	
His	Pro	Arg	11e		Ser	Ser	Arg	Lys 1385		Glu	Leu	Arg	Ala 1390		Leu
Cys	Thr	Asn 139		Leu	Ile	Tyr	Asp 1400		Ala	Pro	Leu	Ile 1405		Arg	Asp
Ala	Thr 141		Leu	Tyr	Thr	Gln 141		His	Arg	Arg	His 1420		Val	Glu	Phe
Val 142		Trp	Ser	Thr	Pro 143		Leu	Tyr	His	11e 143!		Ala	Lys	Ser	Thr 1440
Ala	Leu	Ser	Met	Ile 144		Leu	Val	Thr	Lys 145		Glu	Lys	Asp	His 145	
Yeu	Glu	Ile	Ser 146		Leu	Ile	Gly	Asp 146!	_	Asp	Ile	Asn	Ser 147		Ile
Thr	Glu	Phe		Leu	Ile			Arg		Phe	Thr	Ile 148	-	Leu	Gly

- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565
- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Thr Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635 1640 1645
- Arg Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1660
- Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys 1665 1670 1675 1680
- Gin Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala 1685 1690 1695
- Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn 1700 1705 1710
- Met Ser Ile Lys Asp Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu 1715 1720 1725
- Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly 1730 1735 1740
- Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn 1745 1750 1755 1760

- 104 -

Ser	Ser	Ala	Сув	Tyr	Lys	Ala	Val	Glu	Ile	Ser	Thr	Leu	Ile	Arg	Arg
		176						·	177	0				177	5

- Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly 1780 1785 1790
- Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe 1795 1800 1805
- Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu 1810 1815 1820
- Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val 1825 1830 1835 1840
- Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp

  1845 1850 1855
- Val Gly Ser Val Asp Cys Phe Asn Phe Ile Val Ser Asn Ile Pro Thr 1860 1865 1870
- Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asn Lys 1875 1880 1885
- Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala 1890 1895 1900
- Leu Leu Cly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro 1905 1910 1915 1920
- Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser His 1925 1930 1935
- Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser 1940 1945 1950
- Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965
- Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1975 1980
- Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000
- Ile Gln Ala Ile Val Gly Asp Ala Val Ser Arg Gly Asp Ile Asn Pro 2005 2010 2015
- Thr Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Asn Cys Gly
  2020 2025 2030
- Leu Ala Ile Asn Gly Pro Lys Leu Cys Lys Glu Leu Ile His His Asp

PCT/US97/16718 WO 98/13501

- 105 -

	2035	i		2040					2045	•			
	a Ser 50	Gly Gln	Asp Gly 205		Leu	Asn		Ile 2060		Ile	Leu	Tyr	
Arg Gl 2065	u Leu	Ala Arg	Phe Lys	qaA	Asn	Gln	Arg 2075		Gln	Gln	Gly	Met 2080	
Phe Hi	s Ala	Tyr Pro 208	Val Leu 5	Val	Ser	Ser 2090		Gln	Arg	Glu	Leu 2095		
Ser Ar	g Ile	Thr Arg 2100	Lys Phe	_	Gly 2105		Ile	Leu	Leu	Tyr 2110		Gly	
Asn Ar	g Lys 2115		Asn Lys	Phe 2120		Gln	Asn		Lys 2125		Gly	Tyr	
	e Leu 30	Asp Leu	His Glm 213		Ile	Phe		Lув 2140		Leu	Ser	Lys	
Ser Gl 2145	u Lys	Gln Ile	Ile Met 2150	Thr	Gly	Gly	Leu 2155		Arg	Glu	Trp	Val 2160	
Phe Ly	s Val	Thr Val 216	Lys Glu 5	Thr	Lys	Glu 2170	-	Tyr	Lys	Leu	Val 2175	_	
Tyr Se	r Ala	Leu Ile 2180	Lys Asp	•									
(2) INFORMA	TION F	OR SEQ	ID NO:3:										
(	(2) INFORMATION FOR SEQ ID NO:3:  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 15894 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: DNA (genomic)												
(xi) SE	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:												
ACCAAACAAA	GTTGGG	STAAG GA	TAGATCAA	TCAA	TGAT	CA T	ATTC	TAGT	A CA	CTTA	.GGAT		6
TCAAGATCCT	TCAAGATCCT ATTATCAGGG ACAAGAGCAG GATTAGGGAT ATCCGAGATG GCCACACTTC 12												
TAAGGAGCTT	TAAGGAGCTT AGCATTGTTC AAAAGAAACA AGGACAAACC ACCCATTACA TCAGGATCCG 18												

240

GTGGAGCCAT CAGAGGAATC AAACACATTA TTATAGTACC AATCCCGGGA GATTCCTCAA

- 106 -

TTACCACTC	ATCTAGACTT	CTGGACCGGT	TGGTCAGGTT	AATTGGAAAC	CCGGATGTGA	300
GCGGGCCCAI	ACTAACAGGG	GCACTAATAG	GTATATTATC	CTTATTTGTG	GAGTCTCCAG	360
GTCAATTGAT	TCAGAGGATO	ACCGATGACO	CTGACGTTAG	CATAAGGCTG	TTAGAGGTTG	420
TCCAGAGTGA	CCAGTCACAA	TCTGGCCTTA	CCTTCGCATC	AAGAGGTACC	AACATGGAGG	480
ATGAGGCGGA	CCAATATTT	TCACATGATG	ATCCAAGTAG	TAGTGATCAA	TCCAGGTTCG	540
GATGGTTCGA	GAACAAGGAA	ATCTCAGATA	TTGAAGTGCA	AGACCCTGAG	GGATTCAACA	600
TGATTCTGGG	TACCATCCTA	GCTCAAATTT	GGGTCTTGCT	CGCAAAGGCG	GTTACGGCCC	660
CAGACACGGC	AGCTGATTCG	GAGCTAAGAA	GGTGGATAAA	GTACACCCAA	CAAAGAAGGG	720
TAGTTGGTGA	ATTTAGATTG	GAGAGAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC	CTTACGCCGA	TTCATGGTCG	CTCTAATCCT	GGATATCAAG	AGAACACCCG	840
GGAACAAACC	CAGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATC	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCCTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGGT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAT	CTTTACCAGC	1020
AAATGGGGGA	AACTGCACCA	TACATGGTAA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATC	ATACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG	AGGTTTGAAC	TTTGGCCGAT	CTTACTTCGA	TCCAGCATAT	TTCAGACTAG	1200
GGCAAGAGAT	GGTGAGGAGG	TCAGCTGGAA	AGGTCAGTTC	CACATTGGCA	TCTGAACTCG	1260
GTATCACTGC	CGAAGATGCA	AGGCTTGTTT	CAGAGATCGC	AATGCATACT	ACAGAGGACA	1320
GGATCAGTAG	AGCGGTTGGA	CCCAGACAAT	CCCAAGTGTC	ATTCCTACAC	GGTGATCAAA	1380
ATGAAAATGA	GCTACCGAGA	TGGGGGGGTA	AGGAAGATAT	GAGGGTCAAA	CAGAGTCGGG	1440
GAGAAGCCAG	AGAGAGCTAC	AGAGAAACCA	GGCCCAGCAG	AGCAAGTGAC	GCGAGAGCTA	1500
CCCATCCTCC	AACCGACACA	CCCTTAGACA	TTGACACTGC	ATCGGAGTCC	AGCCAAGATC	1560
CGCAGGACAG	TCGAAGGTCA	GCTGACGCCC	TGCTCAGGCT	GCAAGCCATG	GCAGGAATCT	1620
CGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTAGAGTGTA	CAATGACAGA	GATCTTCTAG	1680
actaggtgca	AGAGGCCGAG	GACCAGAACA	ACATCCGCCT	ACCCTCCATC	attgttataa	1740
AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCCACCA	ACCATCCACT	CCCACGATTG	1800

- 107 -

GGGCCGATGG CAGAAGAGCA GGCACGCCAT GTCAAAAACG GACTGGAATG CATCCGGGCT	1860
CTCAAGGCCG AGCCCATCGG CTCACTGGCC ATCGAGGAAG CTATGGCAGC ATGGTCAGAA	1920
ATATCAGACA ACCCAGGACA GGAGCGAGCC GCCTGCAAGG AAGAGAAGGC AAGCAGTCCG	1980
GGTCTCAGCA AACCATGCCT CTCAGCAATT GGATCAACTG AAGGCGGTGC ACCTCGCATC	2040
CGCGGTCAGG GATCTGGAGA GAGCGATGAC GACGCTGAAA CTTTGGGAAT CCCCTCAGGA	2100
AATCTCCAGG CATCAAGCAC TGGGTTACAG TGTTATTATG TTTATGATCA CAGCGGTGAA	2160
GCGGTTAAGG GAATCCAAGA TGCTGACTCT ATCATGGTTC AATCAGGCCT TGATGGTGAT	2220
AGCACCCTCT CAGGAGGAGA CAATGAATCT GAAAACAGCG ATGTGGATAT TGGCGAACCT	2280
GATACCGAGG GATATGCTAT CACTGACCGG GGATCTGCTC CCATCTCTAT GGGGTTCAGG	2340
GCTTCTGATG TTGAAACTGC AGAAGGAGGG GAGATCCACG AGCTCCTGAG ACTCCAATCC	2400
AGAGGCAACA ACTITCCAAA GCTTAGGAAA ACTCTCAATG TTCCCCCGCC CCCGGACCCT	2460
GGTAGGGCCA GCACTTCCGA GACACCCATT AAAAAGGGCA CAGACGCGAG ATTAGCCTCA	2520
TTTGGAACGG AGATCGCGTC TTTATTGACA GGTGGTGCAA CCCAATGTGC TCGAAAGTCA	2580
CCCTCGGAAC CATCAGGGCC AGGTGCACCT GCGGGGAATG TCCCCGAGTG TGTGAGCAAT	2640
GCCGTACTGA TACAGGAGTG GACACCCGAA TCTGGTACCA CAATCTCCCC GAGATCCCAG	2700
AATAATGAAG AAGGGGGAGA TTATTATGAT GATGAGCTGT TCTCTGATGT CCAAGATATT	2760
AAAACAGCCT TGGCCAAAAT ACACGAGGAT AATCAGAAGA TAATCACCAA GCTAGAATCA	2820
CTGCTGTTAT TGAAGGGGGA AGTTGAGTCA ATCAAGAAGC AGATCAACAG GCAAAATATC	2880
AGCATATCCA CCTTGGAAGG ACACCTCTCA AGCATCATGA TCGCCATTCC TGGACTTGGG	2940
AAGGATCCCA ACGACCCCAC TGCAGATGTC GAAATCAATC CCGACTTGAA ACCCATCATA	3000
GGCAGAGATT CAGGCCGAGC ACTGGCTGAA GTTCTCAAGA AACCCGTTGC CAGCCGACAA	3060
ATCCAAGGAA TGACAAATGG ACGGACCAGT TCCAGAGGAC AGCTGCTGAA GGAATTTCAG	3120
CTAAAGCCGA TCGGGAAAAA GATGAGCTCA GCCGTCGGGT TTGTTCCGGA CACCGGCCCT	3180
GCATCACGCA GTGTAATCCG CTCCATTATA AAATCCAGCC GGCTAGAGGA GGATCGGAAG	3240
CGTTACCTGA TGACTCTCCT TGATGACATC AAAGGAGCCA ACGATCTTGC CAAGTTCCAC	3300
CAGATGCTGA TGAAGATAAT AATGAAGTAG CTACAGCTCA ACTTACCTGC CAACCCCATG	3360

- 108 -

CCAGTCGACC	TAGCTAATAC	AACCTAAATC	CATTATAAAA	AACTTAGGAG	CAAAGTGATT	3420
GCCTCCCAAG	TTCCACAATG	ACAGAGATCT	ACGACTTCGA	CAAGTCGGCA	TGGGACATCA	3480
AAGGGTCGAT	CGCTCCGATA	CAACCCACCA	CCTACAGTGA	TGGCAGGCTG	GTGCCCCAGG	3540
TCAGAGTCAT	AGATCCTGGT	CTAGGCGACA	GAAAAGATGA	ATGTTTTATG	TACATGTTTC	3600
TGCTGGGGGT	TGTTGAGGAC	AGCGATCTCC	TAGGGCCTCC	AATCGGGCGA	GCATTTGGGT	3660
CTCTGCCCTT	AGGTGTTGGC	AGATCCACAG	CAAAACCCGA	AGAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTGA	CATAGTTGTT	AGACGTACAG	CAGGGCTCAA	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCCC	ACTAACTCTC	CTCATACCTT	GGAGAAAGGT	CCTAACAACA	GGGAGTGTCT	3840
TCAACGCAAA	CCAAGTGTGC	AATGCGGTTA	ATCTGATACC	GCTGGATACC	CCGCAGAGGT	3900
TCCGTGTTGT	TTATATGAGC	ATCACCCGTC	TTTCAGATAA	CGGGTATTAC	ACCGTTCCTA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	GGCGATTGGC	CATGGGAAGA	TCATCGACAA	TGCAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	AAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCCTATGT	TACCCACTGA	TGGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380
AAGAATTCCG	CATTTACGAC	GACGTTATCA	TAAATGATGA	CCAAGGATTA	TTCAAAGTTC	4440
TGTAGACCGT	AGTGCCCAGC	AATGCCCGAA	GACGACCCTC	CTCACAATGA	CAGCCAGAAG	4500
GCCCGGAAAA	AAAGGCCCCC	TCCGAAAGAC	TCCACAGACC	AAATGAGAGG	CCAGCCAGCA	4560
GCTGACGGCA	AGCACGAACA	CCAGGCGGCC	CCAGCACAGA	ACAGCCCTGA	CATAAGGCCA	4620
CCACCAGCCA	TCCCAATCTG	CATCCTCCTC	GTAGGACCCC	CGAGGACCAA	CCCCCAAGGT	4680
TGCCCCCAC	CCAAACCACC	AACCGCATCC	CTACCACCCC	CGGGAAAGAA	ACCCCCAGCA	4740
ACTGGAAGAG	CCCTTCCCCT	TTCCCTCAAC	ACAAGAACTC	CACAACCGAA	CCACACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCACCCGA	CTCCCTAGAC	AGATCCTCTC	CCCCTGGCAA	4860
ACTARACARA	ACTTAGGGCC	AAGGAACATA	CACACCCAAC	AGAACCCAGA	CCCCGGCCCA	4920

- 109 -

CGGCGCCGCG	CCCCCAACCC	CCGACAACCA	GAGGGAGCCC	CCAACCAATO	CCGCCGGCTC	4980
CCCCGGTGCC	CACAGGCAGG	CACACCAACC	CCCGAACAGA	CCCAGCACCC	AGCCATCGAC	5040
AATCCAAGAC	GGGGGGCCC	CCCCAAAAA	AGGCCCCCAG	GGGCCGACAG	CCAGCACCGC	5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CGACAACCAA	ACCAGAACCC	AGACCACCCT	5160
GGGCCACCAG	TTCCCAGACT	CGGCCATCAC	CCCGCAGAAA	GGAAAGGCCA	CAACCTGCGC	5220
ACCCCAGCCC	CGATCCGGCG	GGCAGCCACC	CAACCCTAAC	CAGCACCCAA	GAGCGATCCC	5280
CGAAGGACCC	CCGAACCGCA	AAGGACATCA	GTATCCCACA	GCCTCTCCAA	GTCCCCCGGT	5340
CTCTTCCTCT	TCTCGAAGGG	ACTAAAAGAT	CAATCCACCA	CATCCGACGA	CACTCAACTC	5400
CCCGTCCCTA	AAGGAGACAC	CGGGAATCCC	GGAATTAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGAAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTAG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGTTT	5760
CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCAG	GAGTAGTCCT	GGCAGGTGCG	5820
GCCCTAGGCG	TTGCCACAGC	TGCTCAGATA	ACAGCCGGCA	TTGCACTTCA	CCAGTCCATG	5880
CTGAACTCTC	AAGCCATCGA	CAATCTGAGG	GCAAGTCTGG	AAACTACTAA	TCAGGCAATT	5940
GAGGCAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
ATCAATAATG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTCGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCTAGCTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	CGGAGGAGAT	6180
ATCAATAAGG	TGTTAGAAAA	GCTCGGATAT	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
AGTATAGCCT	ACCCGACGCT	GTCCGAGATC	AAGGGGGTGA	TTGTCCACCG	GCTAGAGGGG	6360
GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACGACTG	TGCCCAAGTA	TGTTGCAACC	6420
CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480

- 110 -

GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGGG	6540
TCCACCAAGT	CCTGTGCTCG	TACACTCGTA	TCTGGGTCTT	TTGGGAACCG	GTTCATTTTG	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGACGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATCATTGG	agaggttgga	CGTAGGGACA	6B40
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	AATTGTTGGA	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTGTCG	AGCACTAGCA	TAGTCTACAT	CCTGATTGCA	6960
GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACAGGA	7080
ACATCGAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGGAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCATCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TCCCCTTTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	AGTAGGATAG	TTATCAACAG	AGAACACCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTCG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CAATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
TCTAGATGTA	ACTAACTCAA	TTGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
CATCTCTGAC	AAGATTAAAT	TCCTTAACCC	GGATAGGGAG	TACGACTTCA	GAGATCTCAC	7680
TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
GGCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
ATTCTCAAAC	ATGTCGCTGT	CCCTGTTGGA	CTTGTATTTA	AGTCGAGGTT	ACAATGTGTC	7920
ATCTATAGTC	ACTATGACAT	CCCAGGGAAT	GTACGGGGGA	ACTTACCTAG	TGGAAAAGCC	7980
TAATCTGAGC	AGCAAAGGGT	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040

- 111 -

AGGTGTTATC AGAAATCCGG GTTTGGGGGC TCCGGTGTTC CATATGACAA ACTATTTTGA	8100
GCAACCAGTC AGTAATGATC TCAGCAACTG TATGGTGGCT TTGGGGGAGC TCAAACTCGC	8160
AGCCCTTTGT CACGGGGGAG ATTCTATCAC AATTCCCTAT CAGGGATCAG GGAAAGGTGT	8220
CAGCTTTCAG CTCGTCAAGC TAGGTGTCTG GAAATCCCCA ACCGACATGC AATCCTGGGT	8280
CCCCTTCTCA ACGGATGACC CAGTGATAGA CAGGCTTTAC CTCTCATCTC ACAGAGGTGT	8340
TATCGCTGAC AATCAAGCAA AATGGGCTAT CCCGACAACA AGAACAGATG ACAAGTTGCG	8400
AATGGAGACA TGCTTCCAGC AGGCGTGTAA GGGTAAAATC CAAGCACTCT GCGAGAATCC	8460
CGAGTGGGCA CCATTGAAGG ATAACAGGAT TCCTTCATAC GGAGTCTTGT CTGTTGATCT	8520
GAGTCTAACA GTTGAGCTTA AAATCAAAAT TGCTTCGGGA TTCGGGCCAT TGATCACACA	8580
CGGTTCAGGG ATGGACCTAT ACAAGTCCAA CCACAACAAT GAGTATTGGC TGACTATCCC	8640
GCCAATGAAG AACCTAGCCC TAGGTGTAAT CAACACATTG GAGTGGATAC CGAGATTCAA	8700
GGTTAGTCCC AACCTCTTCA CTGTCCCAAT TAAGGAAGCA GGCGAAGACT GCCATGCCCC	8760
AACATACCTA CCTGCGGAGG TGGATGGTGA TGTCAAACTC AGTTCCAATC TGGTGATCCT	8820
ACCTGGTCAA GATCTCCAAT ATGTTTTGGC AACCTACGAT ACTTCCAGGG TTGAACATGC	8880
TGTGGTTTAT TACGTTTACA GCCCAAGCCG CTCATTTTCT TACTTTTATC CTTTTAGGTT	8940
GCCTATAAAG GGGATCCCCA TCGAATTACA AGTGGAATGC TTCACATGGG ACCAAAAACT	9000
CTGGTGCCGT CACTTCTGTG TGCTTGCGGA CTCAGAATCT GGTGGACATA TCACTCACTC	9060
TGGGATGGTG GGCATGGGAG TCAGCTGCAC AGTCACCCGG GAAGATGGAA CCAATAGCAG	9120
ATAGGGCTGC CAGTGAACCA ATCACATGAT GTCACCCAGA CATCAGGCAT ACCCACTAGT	9180
GTGAAATAGA CATCAGAATT AAGAAAAACG TAGGGTCCAA GTGGTTCCCC GTTATGGACT	9240
CGCTATCTGT CAACCAGATC TTATACCCCG AAGTTCACCT AGATAGCCCG ATAGTTACCA	9300
ACAAGATAGT AGCCATCCTG GAGTATGCTC GAGTCCCTCA CGCTTACAGC CTGGAGGACC	9360
CTACACTGTG TCAGAACATC AAGCACCGCC TAAAAAACGG ATTTTCCAAC CAAATGATTA	9420
TAAACAATGT GGAAGTTGGG AATGTCATCA AGTCCAAGCT TAGGAGTTAT CCGGCCCACT	9480
CTCATATTCC ATATCCAARC TGTAATCAGG ATTTATTTAA CATAGAAGAC AAAGAGTCAA	9540
CGAGGAAGAT CCGTGAACTC CTCAAAAAGG GAAATTCGCT GTACTCTAAA GTCAGTAATA	9600

- 112 -

AGGTTTTCCA	ATGCTTGAGG	GACACTAATT	CACGGCTTGG	TCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAGTT	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAATGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TAACATTTGA	GCTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CTGCTATGAC	CATTGATGCT	AGATATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AATTGATAGA	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCCATGC	10080
	TTCACTTGCT					10140
	CCACTGCTTT					10200
	TTATCACGAG					10260
	AGGGGAGATT					10320
	TGCTGAAAAT					10380
	GAAAGGTCAT	•				10440
	CAGTTGGCCA					10500
	CTCAGGTGAA					10560
	GAAATTTGGC					10620
	CAAGGCACTT					10680
	TTACGACCCC					10740
	GAGCTTTGAC					10800
	TGAGTTCAAC					10860
	TGCTAAAATG					10920
	CGGGATTGGC					10980
	GGCACTCCAC					11040
					ACAAATACCA	11100
-3-1-C-1-GAG	GGCAGCAAAA	GGGIIIATAG	GGTTCCCTCA	GATAATTCGG	CAGGACCAAG	11160

- 113 -

ACI	CTAATCA	TCCGGAGAAT	ATGGAAGCT	ACGAGACAG	CAGTGCATT	T ATCACAACTG	11220
ATO	TCAAGAA	GTACTGCCTT	AATTGGAGA1	ATGAGACCAT	CAGCTTGTT	GCACAGAGGC	11280
TA	latgagat	TTACGGATTA	CCCTCATTT	TTCAGTGGCT	GCATAAGAG	CTTGAGACCT	11340
CTG	STCCTGTA	TGTAAGTGAC	CCTCATTGCC	CCCCGACCT	TGACGCCCAT	ATCCCGTTAT	11400
GCA	AAGTCCC	CAATGACCAA	ATCTTCATTA	AGTACCCTAT	GGGAGGTAT	GAAGGGTATT	11460
GTC	'agaagct	GTGGACCATC	AGCACCATTO	CCTATTTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAG	TAAGGAT	TGCTTCATTA	GTGCAAGGGG	ACAATCAGAC	CATAGCTGTA	ACAAAAAGGG	11580
TAC	CCAGCAC	ATGGCCTTAC	AACCTTAAGA	AATGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACT	TTGTAAT	TCTTAGGCAA	AGGCTACATG	ACATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAA	TTGTTTC	ATCACATTTT	TTTGTTTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGT	CCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAA	CAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATG	ACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAG	ATTCTGATCT	11940
CTC	TTGGCTT	CACAATCAAT	TCAACCATGA	CCCAGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACA	ACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGGG	ATGAATTATC	12060
TGA	atatgag	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATC	TCAAGAG	AATGATTCTC	GCATCACTGA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180
CAC	AGCAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTG:	FATGTGT	CCAGAGCATC	ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTAA	12300
TCC	ACAGTCC	AAACCCAATG	TTAAAGGGAT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGG	BACTGGC	AGCATTCCTC	ATGGACAGGC	Atattatagt	ACCTAGGGCA	GCTCATGAAA	12420
TCCT	rggatca	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTA	GATACCACAA	12480
AAGO	CCTGAT	TCGAGCCAGC	ATGAGGAAGG	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTC	CAATTA	TGACTATGAA	CAATTCAGAG	CAGGGATGGT	GCTATTAACA	GGAAGAAAGA	12600
GAAA	TGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCGAGAGCC	CTAAGAAGCC	12660
ATAT	GTGGGC	GAGGCTAGCT	CGAGGACGGC	CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720

- 114 -

TAGAATCTAT	GCGAGGCCAC	CTTATTCGGC	GTCATGAGAC	ATGTGTCATC	TGCGAGTGTG	12780
GATCAGTCAA	CTACGGATGG	TTTTTTGTCC	CCTCGGGTTG	CCAACTGGAT	GATATTGACA	12840
AGGAAACATC	ATCCTTGAGA	GTCCCATATA	TTGGTTCTAC	CACTGATGAG	AGAACAGACA	12900
TGAAGCTTGC	CTTCGTAAGA	GCCCCAAGTC	GATCCTTGCG	ATCTGCTGTT	AGAATAGCAA	12960
CAGTGTACTC	ATGGGCTTAC	GGTGATGATG	ATAGCTCTTG	GAACGAAGCC	TGGTTGTTGG	13020
CAAGGCAAAG	GGCTAATGTG	AGCCTGGAGG	agctaagggt	GATCACTCCC	ATCTCAACTT	13080
CGACTAATTT	AGCACATAGG	TTGAGGGATC	GTAGCACTCA	AGTGAAATAC	TCAGGTACAT	13140
CCCTTGTCCG	AGTGGCAAGG	TATACCACAA	TCTCCAACGA	CAATCTCTCA	TTTGTCATAT	13200
CAGATAAGAA	GGTTGATACT	AACTTTATAT	ACCAACAAGG	AATGCTCCTA	GGGTTGGGCG	13260
TTTTAGAAAC	ATTGTTTCGA	CTCGAGAAAG	ATACCGGATC	ATCTAACACG	GTATTACATC	13320
TTCACGTCGA	AACAGATTGT	TGCGTGATCC	CAATGATAGA	TCATCCCAGG	ATACCCAGCT	13380
CTCGCAAGCT	AGAGCTGAGG	GCAGAGCTGT	GTACCAACCC	ATTGATATAT	GATAATGCAC	13440
CTTTAATTGA	CAGAGATGCA	ACAAGGCTAT	ACACCCAGAG	CCATAGGAGG	CACCTTGTAG	13500
AATTTGTTAC	ATGGTCCACA	CCCCAACTAT	ATCACATTCT	AGCTAAGTCC	ACAGCACTAT	13560
CTATGATTGA	CCTGGTAACA	AAATTTGAGA	AGGACCATAT	GAATGAAATT	TCAGCTCTCA	13620
TAGGGGATGA	CGATATCAAT	AGTTTCATAA	CTGAGTTTCT	GCTTATAGAG	CCAAGATTAT	13680
TCACTATCTA	CTTGGGCCAG	TGTGCGGCCA	TCAATTGGGC	ATTTGATGTA	CATTATCATA	13740
GACCATCAGG	GAAATATCAG	ATGGGTGAGC	TGTTGTCATC	GTTCCTTTCT	AGAATGAGCA	13800
AAGGAGTGTT	TAAGGTGCTT	GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
GGCACTGTGG	TATTATAGAG	CCTATCCATG	GTCCTTCACT	TGATGCTCAA	AACTTGCACA	13920
CAACTGTGTG	CAACATGGTT	TACACATGCT	ATATGACCTA	CCTCGACCTG	TTGTTGAATG	13980
aagagttaga	AGAGTTTACA	TTTCTTTTGT	GTGAAAGTGA	CGAGGATGTA	GTACCGGACA	14040
GATTCGACAA	CATCCAGGCA	AAACACTTGT	GTGTTCTGGC	AGATTTGTAC	TGTCAACCAG	14100
GGACCTGCCC	ACCAATTCGA	GGTCTAAGAC	CGGTAGAGAA	ATGTGCAGTT	CTAACCGACC	14160
ATATCAAGGC	GGAGGCTAGG	TTATCTCCAG	CAGGATCTTC	GTGGAACATA	AATCCAATTA	14220
TTGTAGACCA	TTACTCATGC	TCTCTGACTT	ATCTTCGGCG	AGGATCGATC	AAACAGATAA	14280

- 115 -

GATTGAGAGT TGATCCAGGA TTCATTTTCG ACGCCCTCGC TGAGGTAAAT GTCAGTCAGC	14340
CAAAGATCGG CAGCAACAAC ATCTCAAATA TGAGCATCAA GGATTTCAGA CCCCCACACG	14400
ATGATGTTGC AAAATTGCTC AAAGATATCA ATACAAGCAA GCACAATCTT CCCATTTCTG	14460
GGGGCAATCT CGCCAATTAT GAAATCCATG CTTTCCGCAG AATCGGGTTG AACTCATCTG	14520
CTTGCTACAA AGCTGTTGAG ATATCAACAT TAATTAGGAG ATGCCTTGAG CCAGGGGAAG	14580
ACGGCTTATT CTTGGGTGAG GGATCGGGTT CTATGTTGAT CACTTATAAG GAGATACTTA	14640
AACTAAACAA GTGCTTCTAT AATAGTGGGG TCTCTGCCAA TTCTAGATCT GGTCAAAGGG	14700
AATTAGCACC CTATCCCTCC GAAGTTGGCC TTGTCGAACA CAGAATGGGA GTAGGTAATA	14760
TTGTCAAGGT GCTCTTTAAC GGGAGGCCCG AAGTCACATG GGTAGGCAGT GTAGATTGCT	14820
TCAATTACAT AGTTAGTAAT ATCCCTACCT CTAGTGTGGG GTTTATCCAT TCAGATATAG	14880
AGACCTTACC TAACAAAGAT ACTATAGAGA AGCTAGAGGA ATTGGCAGCC ATCTTATCGA	14940
TGGCTCTGCT CCTGGGCAAA ATAGGATCAA TACTGGTGAT TAAGCTTATG CCTTTCAGCG	15000
GGGATTTTGT TCAGGGATTT ATAAGTTATG TAGGGTCTCA TTATAGAGAA GTGAACCTTG	15060
TATACCCCAG ATACAGCAAC TTCATATCTA CTGAATCTTA TTTGGTTATG ACAGATCTCA	15120
AGGCTAACCG GCTAATGAAT CCTGAAAAGA TTAAGCAGCA GATAATTGAA TCATCTGTGC	15180
GGACTTCACC TGGACTTATA GGTCACATCC TATCCATTAA GCAACTAAGC TGCATACAAG	15240
CAATTGTGGG AGACGCAGTT AGTAGAGGTG ATATCAATCC TACTCTGAAA AAACTTACAC	15300
CTATAGAGCA GGTGCTGATC AATTGCGGGT TGGCAATTAA CGGACCTAAA CTGTGCAAAG	15360
AATTGATCCA CCATGATGTT GCCTCAGGGC AAGATGGATT GCTTAATTCT ATACTCATCC	15420
TCTACAGGGA GTTGGCAAGA TTCAAGGACA ACCAAAGAAG TCAACAAGGG ATGTTCCACG	15480
CTTACCCCGT ATTGGTAAGT AGCAGGCAAC GAGAACTTAT ATCTAGAATC ACTCGCAAAT	15540
TTTGGGGGCA CATTCTTCTT TACTCCGGGA ACAGAAAGTT GATAAATAAG TTTATCCAGA	15600
ATCTCAAGTC CGGTTATCTG ATACTAGACT TACACCAGAA TATCTTCGTT AAGAATCTAT	15660
CCAAGTCAGA GAAACAGATT ATTATGACGG GGGGTTTGAA ACGTGAGTGG GTTTTTAAGG	15720
TAACAGTCAA GGAGACCAAG GAATGGTATA AGTTAGTCGG ATACAGTGCC CTGATTAAGG	15780
ACTARTIGGT TGAACTCCGG AACCCTAATC CTGCCCCAGG TGGTTAGGCA TTATTTGTAA	15840

- 116 -

## TATATTAAAG AAAACTTTGA AAATACGAAG TTTCTATTCC CAGCTTTGTC TGGT

15894

- (2) INFORMATION FOR SEQ ID NO:4:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2183 amino acids
      (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
  - Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15
  - Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala 20 25 30
  - Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45
  - Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 60
  - Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80
  - Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn 85 90 95
  - Ile Glu Asp Lys Glu Ser Thr Arg Lys Ile Arg Glu Leu Leu Lys Lys
    100 105 110
  - Gly Asn Ser Leu Tyr Ser Lys Val Ser Asn Lys Val Phe Gln Cys Leu 115 120 125
  - Arg Asp Thr Asn Ser Arg Leu Gly Leu Gly Ser Glu Leu Arg Glu Asp 130 135 140
  - Ile Lys Glu Lys Val Ile Asn Leu Gly Val Tyr Met His Ser Ser Gln 145 150 155 160
  - Trp Phe Glu Pro Phe Leu Phe Trp Phe Thr Val Lys Thr Glu Met Arg
    165 170 175
  - Ser Val Ile Lys Ser Gln Thr His Thr Cys His Arg Arg Arg His Thr 180 185 190

Pro	o Val	1 Pho 19!	Pho 5	Th	Gly	y Se:	200		l Glu	ı Lev	ı Lev	205		: Arg	g As
Let	210		a Ile	Ile	Se:	215		ı Ser	Glr	n His	Va]		Туг	: Let	ı Th
Phe 225	Glu G	l Let	ı Val	Let	230		с Сув	Asp	Val	11e 235		Gly	/ Arg	, Lev	24
The	Glu	Th:	Ala	Met 245		: Ile	э Авр	Ala	Arg 250	Tyr	Thr	: Glu	. Leu	Lev 255	
Arg	y Val	. Arg	260	Met	Tr	Lys	Leu	11e 265		Gly	Phe	Phe	270		Let
Gly	Asn	275	Thr	Туг	Gln	Ile	Val 280		Met	Leu	Glu	Pro 285		Ser	Leu
Ala	290		Gln	Leu	Arg	Asp 295		Thr	Val	Glu	Leu 300	Arg	Gly	Ala	Phe
<b>Leu</b> 305	Asn	His	Сув	Phe	Thr 310		Ile	His	Asp	Val 315	Leu	Asp	Gln	Asn	Gly 320
				325					330					335	
			340					345		Gly			350		
		355					360			Ala		365			
	370					375				Val	380				
385					390					Ile 395					400
				405					410	Leu				415	
			420					425		Ser			430		
		435					440			Phe		445			
Gly	Сув	Phe	Met	Pro	Leu	Ser	Leu	Asp	Ser	Asp	Leu	Thr	Met	Tyr	Leu

455

450

- 118 -

Lys 465	Авр	Lув	Ala	Leu	Ala 470	Ala	Leu	Gln	Arg	Glu 475	Trp	Asp	Ser	Val	Tyr 480
Pro	ГÀв	Glu	Phe	Leu 485	Arg	Tyr	Авр	Pro	Pro 490	Lys	Gly	Thr	Gly	Ser 495	Arg
Arg	Leu	Val	Asp 500	Val	Phe	Leu	Asn	<b>Азр</b> 505	Ser	Ser	Phe	Asp	Pro 510	Tyr	Asp
Met	Ile	Met 515	Tyr	Val	Val	Ser	Gly 520	Ala	Tyr	Leu	His	<b>Авр</b> 525	Pro	Glu	Phe
	Leu 530					535					540				
545	Phe				550					555					560
	Asn			565					570					575	_
	Ala		580					585					590		
	Ser	595					600					605	_	_	
	Leu 610					615					620				
625	Arg				630					635					640
	Gln			645					650					655	
	Ala		660					665					670		
	Glu	675					680					685			
	Pro 690					695					700				
705	Tyr				710					715					720
	Leu			725					730					735	
дīĀ	Gly	Ile	Glu	GIA	Tyr	Сув	Gln	Lys	Leu	Trp	Thr	Ile	Ser	Thr	Ile

- 119 -

			74	D				74	5				75	0	
Pr	э Ту	r Le: 75	и <b>Ту</b> : 5	r Lei	u Ala	a Ala	76		u Se	r Gl	y Va	1 Ar		e Al	a Se
Let	1 Va 77	1 Gl <sub>1</sub> 0	n Gl	A yel	e Ası	775		r Il	e Ål:	a Va	1 Thi 780		a Arç	y Va	l Pr
Se:	Th:	r Trį	Pro	Туг	790	ı Lev	Lye	E Ly	s Tr	795		a Ala	Arg	y Vai	1 Th 80
Arg	j As	р Туз	Phe	805	Ile	e Leu	Arg	g Glr	a Arg 810		Hie	) Asp	Ile	61 <sub>3</sub> 815	
His	Let	ı Lys	820	Asn )	Glu	Thr	Ile	825		: Ser	His	Phe	Phe 830		Ту
Sex	Lyt	835	Ile	Туг	Tyr	. Yab	Gly 840		ı Lev	Val	. Ser	Gln 845		Lev	Ly:
Ser	850	Ala	Arg	Сув	Val	Phe 855	Trp	Sex	Glu	Thr	Ile 860		Asp	Glu	Th:
Arg 865	Ala	Ala	Сув	Ser	Asn 870	Ile	Ala	Thr	Thr	Met 875		Lys	Ser	Ile	G1:
Arg	Gly	Tyr	Asp	Arg 885	Tyr	Leu	Ala	Tyr	Ser 890		Asn	Val	Leu	Lys 895	
Ile	Gln	Gln	Ile 900	Leu	Ile	Ser	Leu	Gly 905		Thr	Ile	Asn	Ser 910	Thr	Met
Thr	Gln	Asp 915	Val	Val	Ile	Pro	Leu 920	Leu	Thr	Asn	Asn	Asp 925	Leu	Leu	Ile
Arg	Met 930	Ala	Leu	Leu	Pro	Ala 935	Pro	Ile	Gly	Gly	Met 940	Asn	Туг	Leu	Asn
Met 945	Ser	Arg	Leu	Phe	Val 950	Arg	Asn	Ile	Gly	Авр 955	Pro	Val	Thr	Ser	Ser 960
Ile	Ala	Asp	Leu	Lys 965	Arg	Met	Ile	Leu	Ala 970	Ser	Leu	Met	Pro	Glu 975	Glu
Thr	Leu	His	Gln 980	Val	Met	Thr	Gln	Gln 985	Pro	Gly	Asp	Ser	Ser 990	Phe	Leu
Asp	Trp	Ala 995	Ser	Asp	Pro		Ser 1000		Asn	Leu	Val	Сув 1005		Gln	Ser
Ile	Thr 101(	Arg )	Leu	Leu	Lys	Asn 1015		Thr	Ala	Arg	Phe 1020		Leu	Ile	His

- 120 -

Ser 102	Pro 5	na <i>A</i>	Pro	Met	Leu 103	Lys O	Gly	Leu	Phe	103	a Asp 35	Asp	Sei	Lys	Glu 104
Glu	двр	Glu	Gly	7 Leu 104	Ala 5	Ala	Phe	Leu	105	As <sub>I</sub>	Arg	Hia	Ile	11e	
Pro	Arg	Ala	Ala 106	His O	Glu	Ile	Leu	106	His 5	Ser	. Val	Thr	Gl <sub>y</sub>		Arg
Glu	Ser	Ile 107	Ala 5	Gly	Met	Leu	Asp 108	Thr 0	Thr	Lys	Gly	Leu 108		Arg	Ala
Ser	Met 109	Arg 0	Lys	Gly	Gly	Leu 109	Thr 5	Ser	Arg	Val	Ile 110		Arg	Leu	Ser
Asn 110	Tyr 5	Asp	Tyr	Glu	Gln 111	Phe 0	Arg	Ala	Gly	Met 111	Val 5	Leu	Leu	Thr	Gly 112
Arg	Lys	Arg	Asn	Val 112	Leu 5	Ile	Asp	Lys	Glu 113	Ser 0	Сув	Ser	Val	Gln 113	
Ala	Arg	Ala	Leu 114	Arg 0	Ser	His	Met	Trp 114	Ala 5	Arg	Leu	Ala	Arg 115		Arg
Pro	Ile	Tyr 115	Gly 5	Leu	Glu	Val	Pro 116	<b>Asp</b>	Val	Leu	Glu	Ser 116		Arg	Gly
His	Leu 117(	Ile )	Arg	Arg	His	Glu 1175	Thr	Сув	Val	Ile	Сув 1180		Сув	Gly	Ser
Val 1185	Asn	Tyr	Gly	Trp	Phe 1190	Phe	Val	Pro	Ser	Gly 119	Су <b>в</b> 5	Gln	Leu	Авр	Asp 1200
Ile	ysb	Lys	Glu	Thr 1205	Ser	Ser	Leu	Arg	Val 121		Tyr	Ile	Gly	Ser 1215	
Thr	Asp	Glu	Arg 1220	Thr	Asp	Met	Lys	Leu 1225		Phe	Val	Arg	Ala 1230		Ser
Arg	Ser	Leu 1235	Arg	Ser	Ala	Val	Arg 1240	Ile	Ala	Thr	Val	Tyr 1245		Trp	Ala
ſyr	Gly 1250	Asp	Asp	Asp	Ser	Ser 1255	Trp	Asn	Glu	Ala	Trp 1260		Leu	Ala	Arg
31n 1265	Arg	Ala	Asn	Val	Ser 1270	Leu	Glu	Glu	Leu	Arg 1275	Val	Ile	Thr	Pro	Ile 1280
Ber	Thr	Ser	Thr	Asn 1285	Leu	Ala	His	Arg	Leu 1290	Arg	Asp	Arg	Ser	Thr 1295	

- 121 -

- Val Lys Tyr Ser Gly Thr Ser Leu Val Arg Val Ala Arg Tyr Thr Thr 1300 1305 1310
- Ile Ser Asn Asp Asn Leu Ser Phe Val Ile Ser Asp Lys Lys Val Asp 1315 1320 1325
- Thr Asn Phe Ile Tyr Gln Gln Gly Met Leu Leu Gly Leu Gly Val Leu 1330 1335 1340
- Glu Thr Leu Phe Arg Leu Glu Lys Asp Thr Gly Ser Ser Asn Thr Val 1345 1350 1355 1360
- Leu His Leu His Val Glu Thr Asp Cys Cys Val Ile Pro Met Ile Asp 1365 1370 1375
- His Pro Arg Ile Pro Ser Ser Arg Lys Leu Glu Leu Arg Ala Glu Leu 1380 1385 1390
- Cys Thr Asn Pro Leu Ile Tyr Asp Asn Ala Pro Leu Ile Asp Arg Asp 1395 1400 1405
- Ala Thr Arg Leu Tyr Thr Gln Ser His Arg Arg His Leu Val Glu Phe 1410 1415 1420
- Val Thr Trp Ser Thr Pro Gln Leu Tyr His Ile Leu Ala Lys Ser Thr 1425 1430 1435 1440
- Ala Leu Ser Met Ile Asp Leu Val Thr Lys Phe Glu Lys Asp His Met 1445 1450 1455
- Asn Glu Ile Ser Ala Leu Ile Gly Asp Asp Asp Ile Asn Ser Phe Ile 1460 1465 1470
- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly 1475 1480 1485
- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565
- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Asn Glu Glu

- 122 -

	13/(	,				15/:	•				1286	,			
Leu 1585	Glu 5	Glu	Phe	Thr	Phe 1590		Leu	Сув	Glu	Ser 159		Glu	Asp	Val	Val 1600
Pro	Asp	Arg	Phe	Asp 1605		Ile	Gln	Ala	Lув 1610		Leu	Сув	Val	Leu 1615	
Ąsp	Leu	Tyr	Сув 1620		Pro	Gly	Thr	Сув 162		Pro	Ile	Arg	Gly 1630		Arg
Pro	Val	Glu 1635		Cys	Ala	Val	Leu 1640		qaƙ	His	Ile	Lys 1645		Glu	Ala
Arg	Leu 1650		Pro	Ala	Gly	Ser 1655		Trp	Asn	Ile	Asn 1660		Ile	Ile	Va1
Asp 1665	His	Tyr	Ser	Сув	Ser 1670		Thr	Tyr	Leu	Arg 1675		Gly	Ser	Ile	Lys 1680
Gln	Ile	Arg	Leu	Arg 1685		Asp	Pro	Gly	Phe 1690		Phe	Авр	Ala	Leu 1695	
Glu	Val	Asn	Val 1700		Gln	Pro	Lys	Ile 1705		Ser	Asn	Asn	Ile 1710		Asn
Met	Ser	Ile 1715		Asp	Phe	Arg	Pro 1720		His	Asp	Asp	Val 1725		Lys	Leu
Leu	<b>L</b> ув 1730		Ile	Asn	Thr	Ser 1735		His	Asn	Leu	Pro 1740		Ser	Gly	Gly
Asn 1745	Leu	Ala	Asn	Tyr	Glu 1750		His	Ala	Phe	Arg 1759	_	Ile	Gly	Leu	Asn 1760
Ser	Ser	Ala	Сув	Tyr 1765		Ala	Val	Glu	Ile 1770		Thr	Leu	Ile	Arg 1775	_
Сув	Leu	Glu	Pro 1780		Glu	Asp	Gly	Leu 1785		Leu	Gly	Glu	Gly 1790		Gly
Ser	Met	<b>Leu</b> 1795		Thr	Tyr	Lys	Glu 1800		Leu	Lys	Leu	Asn 1805	_	Сув	Phe
Tyr	Asn 1810		Gly	Val	Ser	Ala 1815		Ser	Arg	Ser	Gly 1820		Arg	Glu	Leu
Ala 1825	Pro	Tyr	Pro	Ser	Glu 1830		Gly	Leu	Val	Glu 1835		Arg	Met	Gly	Val 1840
Gly	Asn	Ile	Val	Lys 1845		Leu	Phe	Asn	Gly 1850		Pro	Glu	Val	Thr 1855	_

- 123 -

Val Gly Ser Val Asp Cys Phe Asn Tyr Ile Val Ser Asn Ile Pro Thr 1860 1865 1870

Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asn Lys 1875 1880 1885

Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala 1890 1895 1900

Leu Leu Gly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro 1905 1910 1915 1920

Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser His 1925 1930 1935

Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser 1940 1945 1950

Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965

Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1975 1980

Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000

Ile Gln Ala Ile Val Gly Asp Ala Val Ser Arg Gly Asp Ile Asn Pro 2005 2010 2015

Thr Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Asn Cys Gly 2020 2025 2030

Leu Ala Ile Asn Gly Pro Lys Leu Cys Lys Glu Leu Ile His His Asp 2035 2040 2045

Val Ala Ser Gly Gln Asp Gly Leu Leu Asn Ser Ile Leu Ile Leu Tyr 2050 2055 2060

Arg Glu Leu Ala Arg Phe Lys Asp Asn Gln Arg Ser Gln Gln Gly Met 2065 2070 2075 2080

Phe His Ala Tyr Pro Val Leu Val Ser Ser Arg Gln Arg Glu Leu Ile 2085 2090 2095

Ser Arg Ile Thr Arg Lys Phe Trp Gly His Ile Leu Leu Tyr Ser Gly 2100 2105 2110

Asn Arg Lys Leu Ile Asn Lys Phe Ile Gln Asn Leu Lys Ser Gly Tyr 2115 2120 2125

- 124 -

Leu Ile Leu Asp Leu His Gln Asn Ile Phe Val Lys Asn Leu Ser Lys 2130 2135 2140

Ser Glu Lys Gln Ile Ile Met Thr Gly Gly Leu Lys Arg Glu Trp Val 2145 2150 2155 2160

Phe Lys Val Thr Val Lys Glu Thr Lys Glu Trp Tyr Lys Leu Val Gly 2165 2170 2175

Tyr Ser Ala Leu Ile Lys Asp 2180

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15894 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACCARACARA GTTGGGTARG GRTAGATCAR TCARTGATCA TATTCTAGTA CACTTAGGAT 60 TCAAGATCCT ATTATCAGGG ACAAGAGCAG GATTAGGGAT ATCCGAGATG GCCACACTTT 120 TAAGGAGCTT AGCATTGTTC AAAAGAAACA AGGACAAACC ACCCATTACA TCAGGATCCG 180 GTGGAGCCAT CAGAGGAATC AAACACATTA TTATAGTACC AATCCCTGGA GATTCCTCAA 240 TTACCACTCG ATCCAGACTA CTGGACCGGT TGGTCAGGTT AATTGGAAAC CCGGATGTGA 300 GCGGGCCCAA ACTAACAGGG GCACTAATAG GTATATTATC CTTGTTTGTG GAGTCTCCAG 360 GTCAATTGAT TCAGAGGATC ACCGATGACC CTGACGTTAG CATCAGGCTG TTAGAGGTTG 420 TCCAGAGTGA CCAGTCACAA TCTGGCCTTA CCTTCGCATC AAGAGGTACC AACATGGAGG 480 ATGAGGCGGA CCAATACTTT TCACATGATG ATCCAAGTAG TAGTGATCAA TCCAGGTCCG 540 GATGGTTCGA GAACAAGGAA ATCTCAGATA TTGAAGTGCA AGACCCTGAG GGATTCAACA 600 TGATTCTGGG TACCATTCTA GCCCAAATTT GGGTCTTGCT CGCGAAGGCG GTTACGGCCC 660 CAGACACGGC AGCTGATTCG GAGCTAAGAA GGTGGATAAA GTACACCCAA CAAAGAAGGG 720 TAGTTGGTGA ATTCAGATTG GAGAGAAAAT GGTTGGATGT GGTGAGGAAC AGGATTGCCG 780

- 125 -

AGGACCTCTC	CTTACGCCG	TTCATGGTC	CTCTAATCC1	GGATATCAAG	AGGACACCCG	840
GGAACAAAC	AAGGATTGCT	GAAATGATA	GTGACATTGA	TACATATATO	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCCT	ACTATTAAGI	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGG7	GAGTTATCC	CACTTGAGTC	CTTGATGAAT	CTTTACCAGC	1020
AAATGGGAGA	AACTGCACCC	TACATGGTA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATO	ATACCCCCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGGGTG	GAACTTGAAA	1140
ACTCCATGGG	AGGTTTGAAC	: TTTGGTCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT	GGTGAGGAGG	TCAGCTGGGA	AAGTCAGTTC	CACATTAGCA	TCTGAACTCG	1260
GTATCACTGC	: TGAGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCACACT	ACTGAGGACA	1320
GGACCAGTAG	AGCGGTTGGA	CCCAGACAAG	CCCAAGTGTC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA	GCTACCAGGA	TTGGGGGCA	AGGAAGATAG	GAGGGTCAAA	CAGAGTCGGG	1440
GAGAAGCCAG	GGAGAGCTAC	AGAGAAACCG	GGTCTAGCAG	AGCAAGCGAT	GCGAGAGCTG	1500
CCCATCTTCC	AACCAGCGCA	CCCCTAGACA	TTGACACTGC	ATCGGAGTCA	GGCCAAGATC	1560
CGCAGGACAG	TCGACGGTCA	GCTGACGCCC	TGCTCAGGCT	GCAAGCCATG	GCAGGAATCT	1620
TGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTAGGGTGTA	CAATGACAGA	GATCTTCTAG	1680
ACTAGGTGCG	AGAGGCCGAG	GACCAGAACA	ACATCCGCCT	ACCCTCCATC	ATTGTTATAA	1740
AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCAACCA	ACCATCCACT	CCTACGACTG	1800
GGGCCGATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	GTCGAGGAAG	CCATGGCAGC	ATGGTCACAA	1920
			ACCCGCAAGG			1980
GGTCTCAGCA	AACCATGCCT	CTCAGCAATT	GGATCAACTG	AAGGCAGTGC	ACCTCGCATC	2040
					CCCCTCAAGA	2100
					CAGCGGTGAA	2160
					TGATGGTGAT	2220
	•				TGGCGAACCT	2280
GATACCGAGG	GATATGCTAT	CACTGACCGG	GGATCTGCTC	CCATCTCTAT	GGGGTTCAGG	2340

- 126 -

					ACTCCAATCT	2400
AGAGGCAACA	ACTTCCCGAA	GCTTGGGAAA	ACTCTCAATG	TTCCTCCGCC	CCCGAACCCC	2460
GGTAGGGCCA	GCACTTCCGA	GACACCCATT	AAAAAGGGGA	CAGACGCGAG	ATTAGCCTCA	2520
TTTGGAGCGG	AGATCGCGTC	TTTATTGACA	GGTGGTGCAA	CCCAATGTGC	TCGAAAGTCA	2580
CCCTCGGAAC	CATCAGGGCC	AGGTGCACCT	GTGGGGAATG	TCCCCGAGTG	TGTGAGCAAT	2640
GCCGCACTGA	TACAGGAGTG	GACACCCGAA	TCTGGTACCA	CAATCTCCCC	GAGATCCCAG	2700
AATAATGAAG	AAGGGGGAGA	TTATTATGAT	GATGAGCTGT	TCTCCGATGT	CCAAGACATC	2760
AAAACAGCCT	TGGCCAAAAT	ACACGAGGAT	AATCAGAAGA	TAATCTCCAA	GCTAGAATCA	2820
CTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAAAAGC	AGATCAACAG	GCAAAATATC	2880
AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TCGCCATTCC	TGGACTTOGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTC	GAACTCAATC	CCGACCTGAA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AACCCGTTGC	CAGCCGACAA	3060
CTCCAAGGAA	TGACAAATGG	ACGGACCAGT	TCCAGAGGAC	AGCTGCTGAA	GGAATTTCAA	3120
CTAAAGCCGA	TCGGGAAAAA	GATGAGCTCA	GCCGTCGGGT	TTGTTCCTGA	CACCGGCCCC	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGA	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ACGATCTTGC	CAAGTTCCAC	3300
CAGATGCTGA	TGAAGATAAT	AATGAAGTAG	CTACAGCTCA	ACTTACCTGC	CAACCTCATG	3360
CCARTCGACC	TAATTAGTAC	AGCCTAAATC	CATTATAAAA	AACTTAGGAG	CAAAGTGATT	3420
GCCTCCCAAG	TTCCACAATG	ACAGAGATCT	ACGACTTCGA	CAAGTCGGCA	TGGGACATCA	3480
AAGGGTCGAT	CGCTCCGATA	CAACCTACCA	CCTACAGTGA	TGGCAGGCTG	GTGCCCCAGG	3540
TCAGAGTCAT	AGATCCTGGT	CTAGGCGACA	GGAAGGATGA	ATGCTTTACG	TACATGTTTC	3600
TGCTGGGGGT	TGTTGAGGAC	AGCGATCCCC	TAGGGCCTCC	AATCGGGCGA	GCATTTGGGT	3660
CCCTGCCCTT	AGGTGTTGGT	AGATCCACAG	CAAAACCCGA	AGAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTGA	CATAGTCGTT	AGACGTACAG	CAGGGCTCAA	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCCC	ACTAACTCTC	CTCACACCTT	GGAGAAAGGT	CCTAACAACA	GGGAGTGTCT	3840
TCAACGCAAA	CCAAGTGTGC	AATGCGGTTA	ATCTGATACC	GCTGGATACC	CCGCAGAGGT	3900

- 127 -

<b>サ</b> ククスサスサスサ	TTATATGAGC	ATCACCCCTC	<b>ጥጥ</b> ርርርእሞአአ	CCCCTATTAC	3.CCCTTCCT3	3960
iccoldilar	iminidade	AICACCCGIC	IIICGGAIAA	CGGGIATIAC	ACCGITCCIA	3900
GAAGAATGCT	AGAATTCAGA	TCGGTCAATG	CAGTGGCTTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	AGCGATTGGC	CCTGGGAAGA	TCATCGATAA	TGCAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	GAGTGAAGTC	TACTCTGCTG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAAAA	GACCTTATGT	TACCCACTGA	TGGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCCC	4380
AAGAATTCCG	CATTTACGAC	GACGTGATCA	TAAATGATGA	CCAAGGACTA	TTCAAAGTTC	4440
TGTAGACCGT	AGTGCCCAGC	AATACCCGAA	AACGACCCCC	CTCATAATGA	CAGCCAGAAG	4500
GCCCGGACAA	AAAAGCCCCC	TCCAAAAGAC	TCCACGGACC	AAGTGAGAGG	CCAGCCAGCA	4560
GCTGACGGCA	AGCGTGAACA	CCAGGCGGCC	TGGGCACAGA	ACAGCCCCGA	CACAAGGCAA	4620
CCACCAGCCA	TCCCAATCTG	CGTCCTCCTC	GTGGGACCCC	CGAGGACCAA	CCCCCAAGGT	4680
CGCCCCGAC	CCAGACCACC	AACCGCATCC	CCACAGCCCC	CGGGAAAGAG	ACCCCCAGCA	4740
ACTGGAAGGC	CCCTCCCCCT	TTCCCTCAAC	GCAAGAACTC	CACAACCGAA	CCGCACAAGC	4800
GATCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGATCCTCTC	CCCCGGCAA	4860
ACTARACAAA	ACTTAGGGCC	AAGGAACATA	CACACCCGAC	AGAACCCAGA	CCCCGGCCCA	4920
CGGCGCCGCG	CCCCCACCTC	CCGACAACCA	GAGGGAGCCC	CCAACCAATC	CCGCCGGCTC	4980
CCCCGGTGCC	CACAGGCAGG	CACACCAACC	CTCGAACAGA	CCCAGCACCC	AGCCATCGAC	5040
AATTCAAGAC	GGGGGCCCC	CCCCAAAAAA	AGGCCCCAG	GGGCCGACAG	CCAGCACCGC	5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CAGGAACCGA	ACCAGAATCC	AGACCACCCT	5160
GGGCCACCAG	TTCCCAGACT	CGGCCATCAC	CCCGCAGAAA	GGAAAGGCCA	CAACCCGCGC	5220
ACCCCTGCCC	TGATCCGGTG	GGCGGCCACC	CAACCCGAAC	CAGCACCCAA	GAGCGATCCC	5280
CGAAGGCCC	CCGAACCGCA	AAAGACATCA	GTATCCCACA	GCCTCTCCAA	GTCCCCCGGT	5340
CTCCCCCTCT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CACCCGACGA	CACTCAATTC	5400
CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460

- 128 -

	GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
	ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGGAT	AGGAAGTGCA	5580
	AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CARTCATTAG	TCATAAAATT	AATGCCCAAT	5640
	ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
	ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGGTT	5760
	CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCTG	GAGTTGTCCT	GGCGGGTGCG	5820
	GCCCTAGGCG	TTGCCACAGC	TGCTCAGATA	ACAGCCGGCA	TTGCACTTCA	CCAGTCCATG	5880
	TTGAACTCTC	AAGCCATCGA	CAATCTGAGA	GCGAGCCTGG	AAACTACTAA	TCAGGCAATT	5940
	GAGGCAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
	ATCARTARTG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
	CTAGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CACTATTTGG	CCCCAGCTTA	6120
	CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	TGGAGGAGAT	6180
	ATCAATAAGG	TGTTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
	AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTACTC	6300
	AGTATAGCCT	ATCCGACGCT	GTCCGAGATT	AAGGGGGTGA	TTGTCCACCG	GCTAGAAGGG	6360
	GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420
	CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
	GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGG	6540
	TCCACCAAGT	CCTGTGCTCG	TACACTTGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
	TCACAAGGGA .	ATCTAATAGC	Caattgtgca	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
	ACGATCATTA .	ATCAGGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
	GTGGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGCGGTATCC	GGACGCTGTG	6780
•	IACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATCATTGG	AGAGGTTGGA	CGTAGGGACA	6840
2	natctgggga :	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	aattgttgga	GTCATCGGAC	6900
•	CAGATATTGA (	GGAGTATGAA	AGGTTTATCG	AGCACTAGCA	TAGTTTACAT	CCTGATTGCA	6960
(	STGTGTCTTG (	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020

- 129 -

	AACAAAAAGG	GAGAACAAGT	' TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACAGGA	7080
	ACATCAAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
	CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCATCCA	GCATCGAGCC	CACCTGAAAT	7200
	TGTCTCCGGA	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
	ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ACAACCCCCA	7320
						GACCTTATGT	7380
	TTTGCTGGCT	GTTCTATTCG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
	CATAAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
	TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
	GATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACCGACC	TAGTGAAATT	7620
						GAGATCTCAC	7680
	TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
	GGCTGCTGAA	GAACTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGG	CCAGGGTAAC	7800
	CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
		ATGTCGCTGT					7920
		ACTATGACAT					7980
	TAATCTGAGC	AGTAAAGGGT	CAGAGTTGTC	ACAACTGAGC	ATGCACCGAG	TGTTTGAAGT	8040
		AGAAATCCGG					8100
		AGTAATGATT			·		8160
		CACAGGGAAG					8220
(	CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCTG	GAAATCCCCA	ACCGACATGC	AATCCTGGGT	8280
(	CCCCTATCA	ACGGATGATC	CAGTGATAGA	CAGGCTCTAC	CTCTCATCTC	ACAGAGGCGT	8340
7	FATCGCTGAC .	AATCAAGCAA	AATGGGCTGT	CCCGACAACA	CGGACAGATG	ACAAGTTGCG	8400
		TGCTTCCAGC					8460
						CTGTTAATCT	8520
G	EAGTCTGACA (	GTTGAGCTTA A	AAATCAAAAT	TGCTTCAGGA	TTCGGGCCAT	TGATCACACA	8580

- 130 -

CGGTTCAGGG	ATGGACCTAT	ACAAATCCAA	CCACAACAAT	GTGTATTGGC	TGACTATCCC	8640
GCCAATGAAG	AACCTAGCCT	TAGGTGTAAT	CAACACATTG	GAGTGGATAC	CGAGATTCAA	8700
GGTTAGTCCC	TACCTCTTCA	CTGTTCCAAT	TAAGGAAGCA	GGCGAGGACT	GCCATGCCCC	8760
AACATACCTA	CCTGCGGAGG	TGGATGGTGA	TGTCAAACTC	AGTTCCAATC	TGGTGATTCT	8820
ACCTGGTCAA	GATCTCCAAT	ATGTTTTGGC	AACCTATGAT	ACTTCCAGAG	TTGAACATGC	8880
TGTGGTTTAT	TACGTTTACA	GCCCAAGCCG	CTCATTTTCT	TACTTTTATC	CTTTTAGGTT	8940
GCCTATAAGG	GGGGTCCCCA	TCGAATTACA	AGTGGAATGC	TTCACATGGG	ACCAAAAACT	9000
CTGGTGCCGT	CACTTCTGTG	TGCTTGCGGA	CTCAGAATCT	GGTGGATATA	TCACTCACTC	9060
TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACTCGG	GAAGATGGAA	CCAACCGCAG	9120
ATAGGGCTGC	CAGTGAACCA	ATCACATGAT	GTCACCCAGA	CATCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATT	AAGAAAAACG	TAGGGTCCAA	GTGGTTCCCC	GTTATGGACT	9240
CGCTATCTGT	CAACCAGATC	TTATACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
ATAAGATAGT	AGCTATCCTG	GAGTATGCTC	GAGTCCCTCA	CGCATACAGC	CTGGAGGACC	9360
CTACACTGTG	TCAGAACATC	AAGCACCGCC	TAAAAAACGG	ATTTTCCAAC	CAAATGATTA	9420
TAAACAATGT	GGAAGTTGGG	AATGTCATCA	AGTCCAAGCT	TAGGAGTTAT	CCGACCCACT	9480
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTAA	CATAGAAGAC	AAAGAGTCAA	9540
CAAGGAAGAT	CCGTGAGCTC	CTCAAAAAGG	GAAATTCGCT	GTACTCCAAA	GTCAGTGATA	9600
AGGTTTTCCA	ATGCCTGAGG	GACACTAACT	CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAATT	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAATGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CAGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TGACGTTTGA	ACTGGTCTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGAC	CATTGATGCT	AGGTATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AACTGATAGA	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTACCAAATT	GTAGCCATGC	10080
TGGAGCCTCT	TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	aacagtagaa	CTCAGAGGTG	10140

- 131 -

CTTTCCTT	AA CCACTGO	TTT ACTG	AAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	1020.0
ATGAAGGT.	AC TTATCAT	GAG TTAA	TTGAAG	CCCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTG	AC AGGGGAG	ATT TTCT	CATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACG	SC TGCTGAA	aat gtta	GGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTG	AT GAAAGGT	CAT GCCA	TATTCT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGA	EG CAGTTGG	CCA CCCC	TGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAA	C TTCAGGT	gaa gggt	TAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAAATCTT	10560
TTGCTGGA	GT GAAATTT	GGC TGCT	TTATGC	CTCTTAGCCT	GGATAGTGAT	CTGACAATGT	10620
ACCTAAAG	BA CAAGGCA	CTT GCTG	CTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680
AGTTCCTG	CG TTACGAC	CCT CCCA	AAGGAA	CTGGGTCACG	GAGGCTTGTA	AATGTTTTCC	10740
TTAATGAT:	C GAGCTTT	GAC CCAT	ATGACA	TGATAATGTA	TGTTGTAAGT	GGAGCTTACC	10800
TCCATGAC	CC TGAGTTC	AAC CTGT	CTTACA	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTT:	TT TGCTAAA	ATG ACTT	ACAAAA	TGAGGGCATG	CCAAGTGATT	GCTGAAAATC	10920
TAATCTCAL	LA CGGGATT	GGC AATT	ATTTTA	AGGACAATGG	GATGGCCAAG	GACGAGCACG	10980
ATTTGACT	A GGCACTC	CAC ACTO	TAGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGG	GGGGCCA	GTC TTAA	AAACCC	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100
AGAACGTG	AGCAGCA	AAA GGGT	TTATAG	GATTCCCTCA	TGTAATTCGG	CAGGACCAAG	11160
ACACTGATO	CA TCCGGAG	AAT ATGG	AGGCTT .	ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAG	A GTACTGC	CTT AATT	GGAGAT .	ATGAGACCAT	CAGCTTATTT	GCACAAAGGC	11280
TARATGAG	TTACGGA	TTA CCCT	CATTTT	TCCAGTGGCT	GCATAAGAGG	CTTGAAACCT	11340
CTGTCCTC	A TGTAAGT	GAC CCTC	ATTGCC	CCCCTGACCT	TGACGCCCAT	GTCCCGTTAT	11400
GCAAAGTC	C CAATGAC	CAA ATCT	TCATTA .	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAG	T GTGGACC.	ATC AGCA	CCATTC	CCTATTTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGG	T TGCTTCG	TTA GTGC	AAGGGG :	ACAATCAGAC	CATAGCCGTA	acaaàaaggg	11580
TACCCAGC	C ATGGCCT	TAC AACC	TTAAGA .	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
<b>ACTTTGTA</b>	T TCTTAGG	CAA AGGC	TACATG	ACATAGGCCA	TCACCTCAAG	GCAAATGAGA	11700

- 132 -

CAATTGTCTC	ATCACATTT	TTTGTCTATT	CAAAAGGAA7	ATATTATGAT	r gggctacitg	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GIGTATTCT	GTCAGAGACT	DTAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATO	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCA	ATCCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	' AGTCATACCC	CTCCTCACAA	12000
ACAACGATCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATCGGGGG	ATGAATTATC	12060
TGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	TCATCACTAA	TGCCTGAAGA	GACCCTTCAT	CAAGTAATGA	12180
CACAACAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTGTATGCGT	CCAGAGCATC	ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTGA	12300
TCCATAGTCC	AAACCCAATG	TTAAAAGGGT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGGGACTGGC	GGCATTCCTC	ATGGACAGGC	ATATTATAGT	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATCA	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTA	GATACCACAA	12480
AAGGCCTGAT	TCGAGCCAGC	ATGAGGAAGG	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATTA	TGACTATGAA	CAATTTAGAG	CAGGGATGGT	GCTATTGACA	GGAAGAAAGA	12600
GAAATGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCTAGAGCC	CTAAGAAGCC	12660
ATATGTGGGC	AAGGCTAGCT	CGAGGACGGC	CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720
TAGAATCTAT	GCGAGGCCAC	CTTATTCGGC	GCCATGAGAC	ATGTGTCATC	TGCGAGTGTG	12780
GATCAGTCAA	CTACGGATGG	TTTTTTGTCC	CCTCGGGTTG	CCAACTGGAT	GATATTGACA	12840
AGGAAACATC	ATCCTTGAGA	GTCCCATATA	TTGGTTCTAC	CACTGATGAG	AGAACAGACA	12900
TGAAGCTTGC	CTTCGTAAGA	GCCCCAAGTC	GATCCTTGCG	ATCTGCTGTT	AGAATAGCAA	12960
CAGTGTACTC	ATGGGCTTAT	GGTGATGATG	ATAGCTCTTG	GAACGAAGCC	TGGTTGTTGG	13020
CAAGGCAAAG	GGCCAATGTG	AGCCTGGAGG	AGCTAAGGGT	GATCACTCCC	ATCTCAACTT	13080
CGACTAATTT	AGCGCATAGG	TTGAGGGATC	GTACCACTCA	agtgaaatac	TCAGGTACAT	13140
CCCTTGTCCG	AGTGGCAAGG	TATACCACAA	TCTCCAACGA	CAATCTCTCA	TTTGTCATAT	13200
CAGATAAGAA	GGTTGATACT	AACTTTATAT	ACCAACAGGG	AATGCTTCTA	GGGTTGGGTG	13260

- 133 -

TTTTAGAAAC ATTGTTTC	BA CTCGAGAAA	3 ATACCGGATO	ATCTAACAC	GTATTACATC	13320
TTCACGTCGA AACAGATT	TGCGTGATC	CGATGATAGA	TCATCCCAGG	ATACCCAGCT	13380
CCCGCAAGCT AGAGCTTAG	G GCAGAGCTA	GTACCAACCC	: ATTGATATAT	GATAATGCAC	13440
CTTTAATTGA CAGAGATGO	A ACAAGGCTAT	r acacccagag	CCATAGGAGG	CACCTTGTGG	13500
AATTIGTTAC ATGGTCCAC	CA CCCCAACTAT	ATCACATTTT	AGCTAAGTCC	ACAGCACTAT	13560
CTATGATTGA CCTGGTAAG	CA AAATTTGAGA	AGGACCATAT	GAATGAAATT	TCAGCTCTCA	13620
TAGGGGATGA CGATATCAI	T AGTTTCATAA	CTGAGTTTCT	GCTTATAGAG	CCAAGATTAT	13680
TCACTATCTA CTTGGGCC	G TGTGCAGCCA	TCAATTGGGC	ATTTGATGTA	CATTATCATA	13740
GACCATCAGG GAAATATCA	G ATGGGTGAGC	TGTTGTCTTC	GTTCCTTTCT	AGAATGAGCA	13800
AAGGAGTGTT TAAGGTGCT	T GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
GGCATTGTGG TATTATAGA	G CCTATCCATG	GTCCTTCACT	TGATGCTCAA	AACTTACACA	13920
CAACTGTGTG CAACATGAT					13980
AAGAGTTAGA AGAGTTCAC					14040
GATTCGACAA TATCCAGGC					14100
GGACCTGCCC ACCAATTCG					14160
ATATCAAGGC AGAGGCTAG					14220
TTGTAGACCA TTACTCATG					14280
GATTGAGAGT TGATCCAGG					14340
CAAAGATCGG CAGCAACAA					14400
ATGATGTTGC AAAATTGCT					14460
GGGGTAATCT CGCCAATTA					14520
CTTGCTACAA AGCTGTTGA					14580
ACCIONATION CTTGGGTGA					
AACTAAACAA GTGCTTCTA: AATTAGCACC CTATCCCTC					
TTGTCAAAGT GCTCTTTAAG					
			TORNOGUNGI	GINGNITGCT	14820

- 134 -

TCAATTTCAT	AGTCAGTAAT	ATCCCTACCT	CTAGTGTGGG	GTTTATCCAT	TCAGATATAG	14880
AGACCTTACC	TAACAAAGAT	ACTATAGAGA	AGCTAGAGGA	ATTAGCAGCC	ATCTTATCGA	14940
TGGCTCTGCT	CCTTGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
GGGATTTTGT	TCAGGGATTT	ATAAGTTATG	TAGGGTCTTA	TTATAGAGAA	GTGAACCTTG	15060
TCTACCCTAG	ATACAGCAAC	TTCATATCTA	CTGAATCTTA	TTTAGTCATG	ACAGATCTCA	15120
AAGCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	GATAATTGAA	TCATCTGTGC	15180
GGACTTCACC	TGGACTTATA	GGTCACATCC	TATCCATTAA	GCAACTAAGC	TGCATACAAG	15240
CAATTGTGGG	AGACGCAGTT	AGTAGAGGTG	GTATCAACCC	TATTCTGAAG	AAACTTACAC	15300
CTATAGAGCA	GGTGCTGATC	AATTGCGGGT	TGGCAATTAA	CGGACCTAAA	CTGTGCAAAG	15360
AATTGATCCA	CCATGATGTT	GCCTCAGGGC	AAGATGGATT	GCTTAACTCT	ATACTCATCC	15420
TCTACAGGGA	GTTGGCAAGA	TTCAAAGACA	ACCAAAGAAG	TCAACAAGGG	ATGTTCCATG	15480
CTTACCCCGT	ATTGGTAAGT	AGCAGGCAAC	GAGAACTTAT	ATCTAGGATC	ACCCGCAAAT	15540
TTTGGGGGCA	TATTCTTCTT	TACTCCGGGA	ACAGAAAGTT	GATAAATCGG	TTTATCCAGA	15600
ATCTCAAGTC	CGGTTACCTG	ATACTAGACT	TACACCAGAA	TATCTTCGTT	AAGAATCTAT	15660
CTAAGTCAGA	GAAACAGATT	ATTATGACGG	GGGGTTTAAA	ACGTGAGTGG	GTTTTTAAGG	15720
TAACAATCAA	GGAGACCAAA	GAATGGTATA	AGTTAGTCGG	ATACAGTGCC	CTGATTAAGG	15780
ATTAATTGGT	TGGACTCCGG	GACCCTAATC	CTGCCCTAGG	TAGTTAGGCA	TTATTTGCAA	15840
TATATTAAAG	AAAACTTTGA	AAATACGAAG	TTTCTATTCC	CAGCTTTGTC	TGGT	15894

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2183 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

- 135 -

	Met 1	Asp	Ser	Leu	Ser 5	Val	Asn	Gln	Ile	Leu 10	Тух	Pro	Glu	Val	His 15	Let
	Asp	Ser	Pro	Ile 20	Val	Thr	Asn	Lys	Ile 25	Val	Ala	Ile	Leu	Glu 30	Tyr	Ala
	Arg	Val	Pro 35	His	Ala	Tyr	Ser	Leu 40	Glu	Asp	Pro	Thr	Leu 45	Сув	Gln	Asr
	Ile	Lys 50	His	Arg	Leu	Lys	Asn 55	Gly	Phe	Ser	Asn	Gln 60	Met	Ile	Ile	Asr
	65					70					75			Ser		В0
					85					90				Leu	95	
				100					105					Leu 110		
			115					120					125	Gln		
		130					135					140		Arg		
	145					150					155			Ser		160
					165					170				Glu	175	
				180					185					Arg 190		
			195					200					205	Ser		_
		210					215					220	_	Tyr		
	225					230					235			Arg		240
					245					250				Leu	255	_
				260					265					Pro 270		
4	TA	Asn	Pro	Thr	Tyr	Gln	Ile	Val	Ala	Met	Leu	Glu	Pro	Leu	Ser	Leu

- 136 -

			275					280					285			
A.	la	Tyr 290	Leu	Gln	Leu	Arg	Asp 295	Ile	Thr	Val	Glu	Leu 300	Arg	Gly	Ala	Phe
	ви 05	Asn	His	Сув	Phe	Thr 310	Glu	Ile	His	Asp	Val 315	Leu	yab	Gln	Asn	Gl <sub>3</sub> 320
Ρĺ	he	Ser	yab	Glu	Gly 325	Thr	Tyr	His	Glu	Leu 330	Ile	Glu	Ala	Leu	Авр 335	Туз
I	le	Phe	Ile	Thr 340	Asp	Asp	Ile	His	Leu 345	. Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
			355					360		Glu			365			
		370					375			Lys		380		-		
31	85					390				Gly	395					400
					405					Pro 410					415	
				420	•				425	Ala				430		
			435					440		Ser			445			
		450					455			Ser		460			-	
46	55					470				Arg	475		_			480
					485					Pro 490					495	
				500					505	Ser				510		
			515					520		Tyr			525			
		530					535			Glu		540				
L∈ 54	15	Phe	Ala	Lys	Met	Thr 550	Tyr	Lys	Met	Arg	Ala 555	Сув	Gln	Val	Ile	Ala 560

- 137 -

Glu	ı Ası	1 Lei	ı Ile	56!		ı Gly	/ Ile	Gl <sub>y</sub>	7 Ass 570		r Phe	Lya	a Asp	57!	n Gly
Met	: Ala	Ly:	<b>3 Asg</b> 580		ı His	a Ası	Leu	585		a Ala	. Leu	Hie	590		ı Ala
Val	. Ser	G1 <sub>3</sub> 595	y Val	Pro	Lys	Asp	Leu 600		Glu	ı Ser	His	Arg		Gly	Pro
Val	610	Lys	Thr	His	Ser	Arg		Pro	Val	. His	620		Thr	Lys	Asn
Val 625	Arg	Ala	Ala	Lys	Gly 630		Ile	Gly	Phe	Pro 635		Val	Ile	Arg	Gln 640
Asp	Gln	Авр	Thr	Asp 645		Pro	Glu	<b>A</b> sn	Met 650		Ala	Tyr	Glu	Thr 655	Val
Ser	Ala	Phe	Ile 660		Thr	Asp	Leu	Lys 665		Tyr	Сув	Leu	Asn 670		Arg
Tyr	Glu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685	Ile	Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	Leu 700	Glu	Thr	Ser	Val
<b>Leu</b> 705	Tyr	Val	Ser	Asp	Pro 710	His	Сув	Pro	Pro	<b>А</b> вр 715	Leu	Asp	Ala	His	Val 720
Pro	Leu	Сув	Lys	Val 725	Pro	Asn	Авр	Gln	Ile 730	Phe	Ile	Lys	Туг	Pro 735	Met
Gly	Gly	Ile	Glu 740	Gly	Tyr	Сув	Gln	<b>L</b> ув 7 <b>4</b> 5	Leu	Trp	Thr	Ile	Ser 750	Thr	Ile
Pro	Tyr	Leu 755	Tyr	Leu	Ala	Ala	Tyr 760	Glu	Ser	Gly	Val	Arg 765	Ile	Ala	Ser
Leu	Val 770	Gln	Gly	Авр	Asn	Gln 775	Thr	Ile	Ala	Val	Thr 780	Lys	Arg	Val	Pro
Ser 785	Thr	Trp	Pro	Tyr	Asn 790	Leu	Lys	Lys	Arg	Glu 795	Ala	Ala	Arg	Val	Thr 800
Arg	Двр	Tyr	Phe	Val 805	Ile	Leu	Arg	Gln	Arg 810	Leu	His	Asp	Ile	Gly 815	His
His	Leu	Lys	Ala 820	Asn	Glu	Thr	Ile	Val 825	Ser	Ser	His	Phe	Phe 830	Val	Tyr

- 138 -

Ser	Lys	Gly 835	Ile	Tyr	Tyr	Asp	Gly 840	Leu	Leu	Val	Ser	Gln 845	Ser	Leu	Lys
Ser	11e 850	Ala	Arg	Сув	Val	Phe 855	Trp	Ser	Glu	Thr	Ile 860	Val	Asp	Glu	Thr
Arg 865	Ala	Ala	Сув	Ser	Asn 870	Ile	Ala	Thr	Thr	Met 875	Ala	Lув	Ser	Ile	Glu 880
	Gly			885					890					895	
	Gln		900					905					910		
	Arg	915					920					925			
	Met 930					935					940				
945	Ser				950					955					960
	Ala			965					970					975	
	Leu		980					985					990		
	Trp	995					1000	)				1005	5		
	Thr 1010	)				1015	5				1020	)			
1025					1030	)				1035	5				1040
	Asp			1045	5				1050	)	_			1055	5
			1060	)				1065	5				1070	)	
	Ser	1075	<b>5</b>				1080	)				1085	5		
	Met 1090 Tyr	)				1095	5				1100	)			
VOII	TAT	vaħ	TAL	GIU	GIII	EIIG	ved	VIG	GIÅ	as t	AGT	rag	neu	THE	GTA

- 139 -

1115

1120

1110

1105

'													L		
Arg	Lys	Arg	Asn	Val 112		Ile	Asp	Lys	Glu 113	Ser O	Сув	Ser	Val	Gln 113	
Ala	Arg	Ala	Leu 114	_	Ser	His	Met	Trp 114		Arg	Leu	Ala	Arg 115	_	Arg
Pro	Ile	Tyr 115	_	Leu	Glu	Val	Pro 116	_	Val	Leu	Glu	Ser 116		Arg	Gly
His	Leu 117		Arg	Arg	His	Glu 117		Сув	Val	Ile	Cys 118		Сув	Gly	Ser
Val 118		Tyr	Gly	Trp	Phe 119		Val	Pro	Ser	Gly 119	_	Gln	Leu	Авр	Asp 1200
Ile	Авр	Lys	Glu	Thr 120		Ser	Leu	Arg	Val 121	Pro 0	Tyr	Ile	Gly	Ser 121	
Thr	Авр	Glu	Arg 122		Asp	Met	Lys	Leu 122!		Phe	Val	Arg	Ala 1230		Ser
Arg	Ser	Leu 123		Ser	Ala	Val	Arg 1240		Ala	Thr	Val	Tyr 124!		Trp	Ala
Tyr	Gly 1250		Asp	Asp	Ser	Ser 125!	_	Asn	Glu	Ala	Trp 1260		Leu	Ala	Arg
Gln 1265	_	Ala	Asn	Val	Ser 1270		Glu	Glu	Leu	Arg 1275		Ile	Thr	Pro	Ile 1280
Ser	Thr	Ser	Thr	Asn 128		Ala	His	Arg	Leu 129	Arg 0	Asp	Arg	Thr	Thr 1295	
Val	Lys	Tyr	Ser 1300	_	Thr	Ser	Leu	Val 1305	-	Val	Ala	Arg	Tyr 1310		Thr
Ile	Ser	Asn 1319		Asn	Leu	Ser	Phe 1320		Ile	Ser	Asp	Lys 1325	-	Val	Asp
Thr	Asn 1330		Ile	Tyr	Gln	Gln 1335	_	Met	Leu	Leu	Gly 1340		Gly	Val	Leu
Glu 1345		Leu	Phe	Arg	Leu 1350		Lys	Asp	Thr	Gly 1355		Ser	Asn	Thr	Val 1360
Leu	His	Leu	His	Val 1365		Thr	Авр	Сув	Сув 1370	Val	Ile	Pro	Met	Ile 1375	_
His	Pro	Arg	Ile 1380		Ser	Ser	Arg	Lys 1385		Glu	Leu	Arg	Ala 1390		Leu

- 140 -

- Cys Thr Asn Pro Leu Ile Tyr Asp Asn Ala Pro Leu Ile Asp Arg Asp 1395 1400 1405
- Ala Thr Arg Leu Tyr Thr Gln Ser His Arg Arg His Leu Val Glu Phe 1410 1415 1420
- Val Thr Trp Ser Thr Pro Gln Leu Tyr His Ile Leu Ala Lys Ser Thr 1425 1430 1435 1440
- Ala Leu Ser Met Ile Asp Leu Val Thr Lys Phe Glu Lys Asp His Met
  1445 1450 1455
- Asn Glu Ile Ser Ala Leu Ile Gly Asp Asp Asp Ile Asn Ser Phe Ile 1460 1465 1470
- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly 1475 1480 1485
- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565
- Ile Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Thr Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635 1640 1645
- Arg Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1660

- 141 -

- Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys 1665 1670 1675 1680
- Gln Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala 1685 1690 1695
- Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn 1700 1705 1710
- Met Ser Ile Lys Asp Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu 1715 1720 1725
- Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly
  1730 1740
- Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn 1745 1750 1755 1760
- Ser Ser Ala Cys Tyr Lys Ala Val Glu Ile Ser Thr Leu Ile Arg Arg 1765 1770 1775
- Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly 1780 1785 1790
- Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe 1795 1800 1805
- Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu 1810 1815 1820
- Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val 1825 1830 1835 1840
- Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp
  1845 1850 1855
- Val Gly Ser Val Asp Cys Phe Asn Phe Ile Val Ser Asn Ile Pro Thr 1860 1865 1870
- Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asn Lys 1875 1880 1885
- Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala 1890 1895 1900
- Leu Leu Gly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro 1905 1910 1915 1920
- Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser Tyr 1925 1930 1935
- Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser

- 142 -

1940 1945 1950

Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965

Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1975 1980

Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000

Ile Gln Ala Ile Val Gly Asp Ala Val Ser Arg Gly Gly Ile Asn Pro 2005 2010 2015

Ile Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Asn Cys Gly
2020 2025 2030

Leu Ala Ile Asn Gly Pro Lys Leu Cys Lys Glu Leu Ile His His Asp 2035 2040 2045

Val Ala Ser Gly Gln Asp Gly Leu Leu Asn Ser Ile Leu Ile Leu Tyr 2050 2055 2060

Arg Glu Leu Ala Arg Phe Lys Asp Asn Gln Arg Ser Gln Gln Gly Met 2065 2070 2075 2080

Phe His Ala Tyr Pro Val Leu Val Ser Ser Arg Gln Arg Glu Leu Ile 2085 2090 2095

Ser Arg Ile Thr Arg Lys Phe Trp Gly His Ile Leu Leu Tyr Ser Gly 2100 2105 2110

Asn Arg Lys Leu Ile Asn Arg Phe Ile Gln Asn Leu Lys Ser Gly Tyr 2115 2120 2125

Leu Ile Leu Asp Leu His Gln Asn Ile Phe Val Lys Asn Leu Ser Lys 2130 2135 2140

Ser Glu Lys Gln Ile Ile Met Thr Gly Gly Leu Lys Arg Glu Trp Val 2145 2150 2155 2160

Phe Lys Val Thr Ile Lys Glu Thr Lys Glu Trp Tyr Lys Leu Val Gly 2165 2170 2175

Tyr Ser Ala Leu Ile Lys Asp 2180

#### (2) INFORMATION FOR SEQ ID NO:7:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15894 base pairs
  - (B) TYPE: nucleic acid

- 143 -

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

# (ii) MOLECULE TYPE: RNA (genomic)

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACCAAACAAA	GTTGGGTAAG	GATAGATCAA	TCAATGATCA	TATTCTAGT	CACTTAGGAT	60
TCAAGATCCT	ATTATCAGGG	ACAAGAGCAG	GATTAGGGAT	' ATCCGAGATG	GCCACACTTT	120
TGAGGAGCTT	AGCATTGTTC	AAAAGAAACA	AGGACAAACC	ACCCATTACA	TCAGGATCCG	180
GTGGAGCCAT	CAGAGGAATC	AAACACATTA	TTATAGTACC	AATTCCTGGA	GATTCCTCAA	240
TTACCACTCG	ATCCAGACTA	CTGGACCGGT	TGGTCAGGTT	AATTGGAAAC	CCGGATGTGA	300
GCGGGCCCAA	ACTAACAGGG	GCACTAATAG	GTATATTATC	CTTATTTGTG	GAGTCTCCAG	360
GTCAATTGAT	TCAGAGGATC	ACCGATGACC	CTGACGTTAG	CATCAGGCTG	TTAGAGGTTG	420
TTCAGAGTGA	CCAGTCACAA	TCTGGCCTTA	CCTTCGCATC	AAGAGGTACC	AACATGGAGG	480
ATGAGGCGGA	CCAATACTTT	TCACATGATG	ATCCAAGCAG	TAGTGATCAA	TCCAGGTCCG	540
GATGGTTCGA	GAACAAGGAA	ATCTCAGATA	TTGAAGTGCA	AGATCCTGAG	GGATTCAACA	600
TGATTCTGGG	TACCATTCTA	GCCCAGATCT	GGGTCTTGCT	CGCAAAGGCG	GTTACGGCCC	660
CAGACACGGC	AGCTGATTCG	GAGCTAAGAA	GGTGGATAAA	GTACACCCAA	CAAAGAAGGG	720
TAGTTGGTGA	ATTTAGATTG	GAGAGAAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC	TTTACGCCGA	TTCATGGTGG	CTCTAATCCT	GGATATCAAG	AGGACACCCG	840
GGAACAAACC	TAGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATC	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCTTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGGT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAT	CTTTACCAGC	1020
AAATGGGAGA	AACTGCACCC	TACATGGTÄA	TCCTAGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GCGCAGGATC	ATACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG .	AGGTTTGAAC	TTTGGTCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT	ggtgaggagg	TCAGCTGGAA .	AGGTCAGTTC	CACATTGGCA	TCCGAACTCG	1260

- 144 -

GTATCACTGC	CGAGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCATACT	ACTGAGGACA	1320
GGATCAGTAG	AGCGGTCGGA	CCCAGACAAG	CCCAAGTATC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA	GCTACCAGGA	TTGGGGGGCA	AGGAAGACAG	GAGGGTCAAA	CAGAGTCGGG	1440
GAGAAGCCAG	GGAGAGCTAC	AGAGAAACCG	AGTCCAGCAG	AGCAAGTGAT	GCGAGAGCTG	1500
CCCATCCTCC	AACCAGCATG	CCCCTAGACA	TTGACACTGC	ATCGGAGTCA	GGCCAAGATC	1560
CGCAGGACAG	TCGAAGGTCA	GCTGACGCTC	TGCTCAGGCT	GCAAGCCATG	GCAGGAATCT	1620
TGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTAGGGTATA	CAATGACAGA	GATCTTCTAG	1680
ATTAGGTGCG	AGAGGCCGAG	GACCAGAACA	ACATCCGCCT	ACCCTCCATC	ATTGTTATAA	1740
AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCAACCA	ACCATCCACT	CCCACGACTG	1800
GAGCCGATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	GTCGAGGAAG	CCATGGCAGC	ATGGTCAGAA	1920
ATATCAGACA	ATCCAGGACA	GGACCGAGCC	GCCTGCAAGG	AAGAGGAGGC	AGGCAGTTCG	1980
GGTCTCAGCA	AACCATGCTT	CTCAGCAATT	GGATCAACTG	AAGGCGGTGC	ACCTCGCATC	2040
CGCGGTCAGG	GATCTGGAGA	AAGCGATGAC	GACGCTGAAA	CTTTGGGAAT	CCCCTCAAGA	2100
AATCTCCAGG	CATCAAGCAC	TGGGTTACAG	TGTTATCATG	TTTATGATCA	CAGCGGTGAA	2160
GCGGTTAAGG	GAATCCAAGA	TGCTGACTCT	ATCATGGTTC	AATCAGGCCT	TGATGGTGAT	2220
AGCACCCTCT	CAGGAGGAGA	CGATGAATCT	GAAAACAGCG	ATGTGGATAT	TGGCGAACCT	2280
GATACCGAGG	GATATGCTAT	CACTGACCGG	GGATCTGCTC	CCATCTCTAT	GGGGTTCAGG	2340
GCTTCTGATG	TTGAAACTGC	AGAAGGAGGG	GAGATCCACG	AGCTCCTGAA	ACTCCAATCC	2400
AGAGGCAACA	ACTTTCCGAA	GCTTGGGAAA	ACTCTCAATG	TTCCTCCGCC	CCCGAACCCC	2460
AGTAGGGCCA	GCACTTCCGA	GACACCCATT	AAAAAGGGGA	CAGACGCGAG	ATTGGCCTCA	2520
TTTGGAACGG	AGATCGCGTC	TTTATTGACA	GGTGGTGCAA	CCCAATGTGC	TCGAAAGTCA	2580
CCCTCGGAAC	CGTCAGGGCC	AGATGCACCT	GCGGGGAATG	TCCCCGAGTG	TGTGAGCAAT	2640
GCCGCACTGA	TACAGGAGTG	GACACCCGAA	TCTGGTACCA	CAATCTCCCC	GAGATCCCAG	2700
AATAATGAAG	AAGGGGGAGA	CTATTATGAT	GATGAGCTGT	TCTCCGATGT	CCAAGACATC	2760
AAAACAGCCT	TGGCCAAAAT	ACACGAGGAT	AATCAGAAGA	TAATCTCCAA	GCTAGAATCA	2820

- 145 -

TTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAGAAGC	AGATCAACAG	GCAAAATATC	2880
AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TTGCCATTCC	TGGACTTGGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTC	GAACTCAATC	CCGACCTGAA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AGCCCGTTGC	CAGCCGACAA	3060
CTCCAGGGAA	TGACTAATGG	ACGGACCAGT	TCCAGAGGAC	AGCTGCTGAA	GGAATTTCAA	3120
CTAAAGCCGA	TCGGGAAAAA	GGTGAGCTCA	GCCGTCGGGT	TTGTCCCTGA	CACCGGCCCT	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGT	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ACGATCTTGC	CAAGTTCCAC	3300
CAGATGCTGA	TGAAGATAAT	AATGAAGTAG	CTACAGCTCA	ACTTACCTGC	CAACCCCATG	3360
CCAGTCGACC	TAATTAGTAC	AACCTAAATC	CATTATAAAA	AACTTAGGAG	CAAAGTGATT	3420
GCCTCCTAAG	TTCCACAATG	ACAGAGATCT	ACGACTTCGA	CAAGTCGGCA	TGGGACATCA	3480
AAGGGTCGAT	CGCTCCGATA	CAACCTACCA	CCTACAGTGA	TGGCAGGCTG	GTGCCCCAGG	3540
TCAGAGTCAT	AGATCCTGGT	CTAGGTGATA	GGAAGGATGA	ATGCTTTATG	TACATGTTTC	3600
TGCTGGGGGT	TGTTGAGGAC	AGAGATCCCC	TAGGGCCTCC	AATCGGGCGA	GCATTCGGGT	3660
CCCTGCCCTT	AGGTGTTGGT	AGATCCACAG	CAAAACCCGA	GGAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTGA	CATAGTTGTT	AGACGTACAG	CAGGGCTCAA	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCCC	ACTAACCCTC	CTCACACCTT	GGAGAAAGGT	CCTAACAACA	GGGAGTGTCT	3840
TCAATGCAAA	CCAAGTGTGC	AATGCGGTTA	ATCTAATACC	GCTGGACACC	CCGCAGAGGT	3900
TCCGTGTTGT	TTATATGAGC	ATCACCCGTC	TTTCGGATAA	CGGGTATTAC	ACCGTTCCCA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTA	GTGACCCTCA	4020
GGATTGACAA	GGCGATTGGC	CCTGGGAAGA	TCATCGACAA	TGCAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAGAA	GAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCTTATGT	TACCCACTGA	TGGATATCAA	TGAAGACCTT	AATCGGTTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380

- 146 -

AAGAATTCCG	CATTTACGAC	GACGTGATCA	A TAAATGATGA	CCAAGGACTA	TTCAAAGTTC	4440
TGTAGACCGT	AGTGCCCAGC	AATACCCGAA	AACGACCCC	CTCATAATGA	CAGCCÁGAAG	4500
GCCCGGACAA	AAAAGCCCCC	TCCAAAAGAC	TTCACGGACC	AAGCGAGAGG	CCAGCCAGCA	4560
GCCGACAGCA	AGTGTGGACA	CCAGGCGGCC	: Caagcacaga	ACAGCCCCGA	CACAAGGCCA	4620
CCACCAGCCA	TCCCAATCCG	CGTCCTCCTC	GTAGGACCCC	CGAGGACCAA	CCCCCAAGGT	4680
CGCTCCGGAC	ACAGACCACC	AGCCGCATCC	CCACAGCCCT	CGGGAAAGGA	ACCCCCAGCA	4740
ACTGGAAGGC	CCCTTCCCCC	CTCCCCCAAC	GCAAGAACCC	CACAACCGAA	CCGCACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGACCCTCCC	TCCCCGGCAT	4860
ACTANACANA	ACTTAGGGCC	AAGGAACACA	CACACCCGAC	AGAACCCAGA	ccccccccc	4920
CGGCACCGCG	CCCCCACCCC	CCGAAAACCA	GAGGGAGCCC	CCAACCAATC	ccgccgcccc	4980
CCCCGGTGCC	CACAGGTAGG	CACACCAACC	CCCGAACAGA	CCCAGCACCC	AGCCACCGAC	5040
AATCCAAGAC	GGGGGGCCCC	CCCCAAAAA	AGGCCCCCAG	GGGCCGACAG	CCAGCATCGC	5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CGGCAACCAA	ACCAGAGCCC	AGACCACCCT	5160
GGGCCACCAG	CTCCCAGACT	CGGCCATCAC	CCCGAAAAA	GGAAAGGCCA	CAACCCGCGC	5220
ACCCCAGGCC	CGATCCGGCG	GGAAGCCACC	CAACCCGAAC	CAGCACCCAA	GAGCGATCCC	5280
TGGGGGACCC	CCAAACCGCA	AAAGACATCA	GTATCCCACC	GCCTCTCCAA	GTCCCCCGGT	5340
CTCCTCCTCT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CATCCGACGA	CACTCAATTC	5400
CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAATGTCTT	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGGAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTGG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAAATTGCAG	AATACAGGAG	ACTACTGAGA	5700
ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGGTT	5760
CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCGG	GAGTTGTCCT	GGCAGGTGCG	5820
	TTGCCACAGC					5880
CTGAACTCTC	AAGCCATCGA	CAATCTGAGA	GCAAGCCTGG	AAACTACTAA	TCAGGCAATT	5940

- 147 -

GAGGCAATC	GGCAAGCAGG	GCAGGAGATO	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
ATCAATAAT	G AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTAGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCCAGCTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCCATC	CAGGCTTTGA	GCTATGCGCT	TGGGGGAGAT	6180
ATCAATAAGG	TATTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
AGTATAGCCT	ATCCGACGCT	GTCCGAGATT	AAGGGGGTGA	TTGTCCACCG	GCTAGAGGGG	6360
GTCTCGTACA	ATATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420
CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGG	6540
TCCACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTCT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTGGTCGAGG	TGAACGGTGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	GGACGCGGTG	6780
			ATATCATTGG			6840
			GATGCCAAGG			6900
			AGCACTAGCA			6960
			GCTTTAATAT			7020
			AGACCAGGCC			7080
			TCCCCTACAA			7140
			ACCGCATCCA			7200
					AGGGTGCAAG	7260
					ACAACCCCCA	7320
					GACCTTATGT	7380
			GAGCTTGATC			7440
CALTAGACTC	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAGAGCC	TCAGCACCAA	7500

- 148 -

TCTAGATGI	A ACTAACTCA	A TCGAGCATC	A GGTCAAGGA	C GTGCTGAC	C CACTCTTCAA	7560
GATCATCGG	T GATGAAGTG	G GCCTGAGGA	C ACCTCAGAG	A TTCACTGAC	C TAGTGAAATT	7620
CATCTCTGA	C AAAATTAAA	T TCCTTAATC	C GGATAGGGA	G TACGACTIO	A GAGATCTCAC	7680
TTGGTGTAT	C AACCCGCCA	g agagaatca	A ATTGGATTA	T GATCARTAC	T GTGCAGATGT	7740
GGCTGCTGA	A GAACTCATG	A ATGCATTGG	T GAACTCAAC	T CTACTGGAG	G CCAGGGCAAC	7800
CAATCAGTT	C CTAGCTGTC	T CAAAGGGAA	A CTGCTCAGG	G CCCACTACA	A TCAGAGGTCA	7860
					T ACAATGTGTC	7920
					G TGGGAAAGCC	7980
					G TGTTTGAAGT	8040
					A ACTATITIGA	8100
					C TCAGGTTCGC	8160
					G GGAAAGGTGT	8220
					AATCCTGGGT	8280
					ACAGAGGTGT	8340
					ACAAGTTGCG	8400
					GCGAGAATCC	8460
					CTGTTAATCT	8520
					TGATCACACA	8580
					TGACTATCCC	8640
					CGAGATTCAA	8700
					GCCATGCCCC	8760
					TGGTAATTCT	8820
					TTGAACATGC	8880
					CTTTTAGGTT	8940
					ACCARARACT	9000
					TCACTCACTC	9060

- 149 -

TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACTCGG	GAAGATGGAA	CCAATCGCAG	9120
ATAGGGCTGC	CAGTGAACCG	ATCACATGAT	GTCACTCAGA	CACCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATT	AAGAAAAACG	TAGGGTCCAA	GTGGTTTCCC	GTCATGGACT	9240
CGCTATCTGT	CAACCAGATC	TTGTACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
ATAAGATAGT	AGCTATCCTG	GAGTATGCTC	GAGTCCCTCA	CGCTTACAGC	CTTGAGGACC	9360
CTACACTGTG	TCAGAACATC	AAGCACCGCC	TAAAAAACGG	ATTCTCCAAC	CAAATGATTA	9420
TAAACAATGT	ggaagttggg	AATGTCATCA	AGTCCAAGCT	TAGGAGTTAT	CCGGCCCACT	9480
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTTAA	CATAGAAGAC	AAAGAGTCAA	9540
CAAGGAAGAT	CCGTGAGCTC	CTAAAAAAGG	GAAATTCGCT	GTACTCCAAA	GTCAGTGATA	9600
AGGTTTTCCA	ATGCCTGAGG	GACACTAACT	CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAATT	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAATGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGC	9840
TGTTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAGGA	GTCTCAACAT	GTATATTACC	9900
TGACGTTTGA	ACTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGAC	CATTGATGCT	AGGTATGCAG	AACTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
aactgataga	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCTATGC	10080
TGGAGCCACT	TTCACTTGCT	TACCTGCAAC	TGAGGGACAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTAA	CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
ATGAAGGTAC	TTATCATGAG	TTAATTGAAG	CCTTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGAC	AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGGC	TGCTGAAAAT	GTCAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGAT	GAAGGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAGG	CAGTTGGCCA	CCCCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAAGC	TTCAGGTGAA	GGGTTAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAGATCAT	10560
TTGCTGGAGT	GAGATTTGGC	TGTTTTATGC	CTCTTAGCCT	GGACAGTGAT	CTGACAATGT	10620

- 150 -

ACCTAAAGGA	CAAGGCACTT	GCTGCTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680
AGTTCCTGCG	TTACGATCCT	CCCAAGGGAA	CCGGGTCACG	GAGGCTTGTA	GATGTTTTCC	10740
TTAATGATTC	GAGCTTTGAC	CCATATGATA	TGATAATGTA	TGTCGTAAGT	GGAGCCTACC	10800
TCCATGACCC	TGAGTTCAAT	CTGTCTTACA	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTTTT	CGCTAAAATG	ACTTACAAAA	TGAGGGCATG	CCAAGTGATC	GCTGAAAATC	10920
TAATCTCAAA	CGGGATTGGC	AAGTATTTTA	AGGACAATGG	GATGGCCAAG	GATGAGCACG	10980
ATTTGACTAA	GGCACTCCAC	ACTCTGGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGGGG	GGGGCCAGTC	TTAAAAACCT	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100
GGAACGTTAA	AGCAGAAAA	GGGTTTGTAG	GATTCCCTCA	TGTAATTCGG	CAGAATCAAG	11160
ACACTGATCA	TCCGGAGAAT	ATAGAAACCT	ACGAGACAGT	CAGCGCATTT	ATCACGACTG	11220
ATCTCAAGAA	GTACTGCCTT	AATTGGAGAT	ATGAGACCAT	CAGCTTATTT	GCACAGAGGC	11280
TAAATGAGAT	TTACGGATTA	CCCTCATTTT	TTCAGTGGCT	GCATAAGAGG	CTTGAAACCT	11340
CTGTCCTCTA	TGTAAGTGAT	CCTCATTGCC	CCCCGACCT	TGACGCCCAT	GTCCCGTTAT	11400
GCAAAGTCCC	CAATGACCAA	ATCTTCATCA	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTACTTATA	CCTGGCTGCT	TATGAGAGCG	11520
GGGTAAGGAT	TGCCTCGTTA	GTGCAAGGGG	ACAATCAGAC	CATAGCCGTA	ACAAAAAGGG	11580
TACCCAGCAC	ATGGCCTTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ACATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
	ACTCAAGAGC					11820
	AGCATGCAGT					11880
					ATTTTGATCT	11940
					CTCCTCACAA	12000
	CTTAATAAGG					12060
					TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCATCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180

- 151 -

CACAACAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTGTATGCGT	CCAGAGCATO	ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTAA	12300
TCCATAGTCC	AAACCCAATG	TTAAAAGGGT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGAGACTGGC	GGCATTCCTC	ATGGACAGGC	ATATTATAGT	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATCA	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTA	GATACCACAA	12480
AAGGCCTGAT	TCGAGCCAGC	ATGAGGAAGG	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATTA	TGACTATGAA	CAATTTAGAG	CAGGGATGGT	GCTATTGACA	GGAAGAAAGA	12600
GAAATGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCTAGAGCC	CTAAGAAGCC	12660
ATATGTGGGC	AAGACTAGCT	CGAGGACGGC	CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720
TAGAATCTAT	GCGAGGCCAC	CTTATTCGGC	GTCATGAGAC	ATGTGTCATC	TGCGAGTGTG	12780
GATCAGTCAA	CTACGGATGG	TTTTTTGTCC	CCTCGGGTTG	CCAACTGGAT	GATATTGACA	12840
AGGAAACATC	ATCCTTGAGA	GTCCCATATA	TTGGTTCTAC	CACTGATGAG	AGAACAGACA	12900
TGAAGCTTGC	CTTCGTAAGA	GCCCCAAGTA	GATCCTTGCG	ATCTGCCGTT	AGAATAGCAA	12960
		GGTGATGATG				13020
CAAGGCAAAG	GGCCAATGTG	AGCCTGGAGG	AGCTAAGGGT	GATCACTCCC	ATCTCGACTT	13080
		TTGAGGGATC				13140
		TATACCACAA				13200
		AACTTTATAT				13260
		CTCGAGAAAG				13320
		TGCGTGATCC				13380
		GCAGAGCTAT				13440
CTTTAATTGA						13500
AATTTGTTAC						13560
					TCAGCTCTCA	13620
TAGGGGATGA						13680
LCACCATCTA	CITGGGCCAG	TGTGCAGCCA	TCAATTGGGC	ATTTGATGTA	CATTATCATA	13740

- 152 -

GACCATCAGG	GAAATATCAG	ATGGGTGAGC	TGTTGTCTTC	GTTCCTTTCT	AGAATGAGCA	13800
AAGGAGTGTT	TAAGGTGCTT	GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
GGCATTGTGG	TATTATAGAG	CCTATCCATG	GTCCTTCACT	TGATGCTCAA	AACTTGCACA	13920
CAACTGTGTG	CAACATGGTT	TACACATGCT	ATATGACCTA	CCTTGACCTG	TTGTTGAATG	13980
AAGAGTTAGA	AGAGTTCACA	TTTCTTTTGT	GTGAAAGCGA	TGAGGATGTA	GTACCGGACA	14040
GATTCGACAA	CATCCAGGCA	AAACACTTGT	GTGTTCTGGC	AGATTTGTAC	TGTCAACCAG	14100
GGACCTGCCC	ACCGATTCGA	GGTCTAAGGC	CGGTAGAGAA	ATGTGCAGTT	CTAACCGATC	14160
ATATCAAGGC	AGAGGCTAGG	TTATCTCCAG	CAGGATCTTC	GTGGAACATA	AATCCAATTA	14220
TTGTAGACCA	TTACTCATGC	TCTCTGACTT	ATCTCCGTCG	AGGATCTATC	AAACAGATAA	14280
GATTGAGAGT	TGATCCAGGA	TTCATTTTTG	ATGCCCTCGC	TGAGGTAAAT	GTCAGTCAGC	14340
CAAAGGTCGG	CAGCAACAAC	ATCTCAAATA	TGAGCATCAA	GGATTTCAGA	CCTCCACACG	14400
ATGATGTTGC	AAAATTGCTC	AAAGATATCA	ACACAAGCAA	GCACAATCTT	CCCATTTCAG	14460
GGGGTAGTCT	TGCCAATTAT	GAAATCCATG	CTTTCCGCAG	AATCGGGTTA	AACTCATCTG	14520
CTTGCTACAA	AGCTGTTGAG	ATATCAACAT	TAATTAGGAG	ATGCCTTGAG	CCAGGGGAAG	14580
ACGGCTTGTT	CTTGGGTGAG	GGGTCGGGTT	CTATGTTGAT	CACTTATAAG	GAGATACTAA	14640
AACTAAACAA	GTGCTTCTAT	AATAGTGGGG	TTTCCGCCAA	TTCTAGATCT	GGTCAAAGGG	14700
AATTAGCACC	CTATCCCTCC	GAAGTTGGCC	TTGTCGAACA	CAGAATGGGA	GTAGGTAATA	14760
TTGTCAAGGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	ATAGATTGCT	14820
TCAATTTCAT	AGTCAGTAAT	ATCCCTACCT	CTAGTGTGGG	ATTTATCCAT	TCAGATATAG	14880
AGACCTTACC	CAACAAAGAT	ACTATAGAGA	agttagagga	ATTGGCAGCC	ATCTTATCGA	14940
TGGCTCTACT	CCTTGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
GGGATTTTGT	TCAGGGATTT	ATAAGCTATG	TAGGGTCTCA	TTATAGAGAA	GTGAACCTTG	15060
TCTACCCTAG	GTACAGCAAC	TTCATATCTA	CTGAATCTTA	TTTAGTTATG	ACAGATCTCA	15120
AAGCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	Gataattgaa	TCATCTGTGC	15180
GGACTTCACC	TGGACTTATA	GGTCACATCC	TATCTATCAA	GCAACTAAGC	TGCATACAAG	15240
CAATTGTGGG	AGGCGCAGTT	agtagaggtg	ATATCAACCC	TATTCTGAAA	AAACTTACAC	15300

- 153 -

CTATAGAGCA GGTGCTGATC AGTTGCGGGT TGGCAATTAA CGGACCTAAG CTGTGCAAAG 15360 AATTAATCCA CCATGATGTT GCCTCAGGGC AAGATGGATT GCTTAACTCT ATACTCATCC 15420 ° TCTACAGGGA GTTGGCAAGA TTCAAAGACA ACCAAAGAAG TCAACAAGGG ATGTTCCACG 15480 CTTACCCCGT ATTGGTAAGT AGTAGGCAAC GAGAACTTGT ATCTAGGATC ACTCGCAAAT 15540 TTTGGGGGCA TATTCTTCTT TACTCCGGGA ACAGAAAGTT GATAAATCGG TTTATCCAGA 15600 ATCTCAAGTC CGGTTATCTA ATACTAGACT TACACCAGAA TATCTTCGTT AAGAATCTAT 15660 CCAAGTCAGA GAAACAGATT ATTATGACGG GGGGTTTAAA ACGTGAGTGG GTTTTTAAGG 15720 TAACAGTCAA GGAGACCAAA GAATGGTATA AGTTAGTCGG ATACAGCGCT CTGATTAAGG 15780 ATTANTIGGT TGAACTCCGG AACCCTAATC CTACCCTAGG TAGTTAGGCA TTATTTGCAA 15840 TATATTAAAG AAAACTTTGA AAATACGAAG TTTCTATTCC CAGCTTTGTC TGGT 15894

### (2) INFORMATION FOR SEQ ID NO:8:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2183 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: protein

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15

Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala 20 25 30

Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45

Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 55 60

Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80

Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn 85 90 95

- 154 -

Ile	Glu	Asp	Lys 100	Glu	Ser	Thr	Arg	Lys 105	Ile	Arg	'Glu	Leu	Leu 110	Lув	Lys
Gly	Asn	Ser 115	Leu	Tyr	Ser	Lys	Val 120	Ser	Asp	Lys	Val	Phe 125	Gln	Сув	Leu
Arg	Asp 130	Thr	Asn	Ser	Arg	Leu 135	Gly	Leu	Gly	Ser	Glu 140	Leu	Arg	Glu	qaA
Ile 145	Lys	Glu	Lys	Ile	Ile 150	Asn	Leu	Gly	Val	Tyr 155	Met	His	Ser	Ser	Gln 160
Trp	Phe	Glu	Pro	Phe 165	Leu	Phe	Trp	Phe	Thr 170	Val	Lys	Thr	Glu	Met 175	Arg
Ser	Val	Ile	Lys 180	Ser	Gln	Thr	His	Thr 185	Сув	His	Arg	Arg	Arg 190	His	Thr
Pro	Val	Phe 195	Phe	Thr	Gly	Ser	Ser 200	Val	Glu	Leu	Leu	11e 205	Ser	Arg	Хsр
Leu	Val 210	Ala	Ile	Ile	Ser	Lys 215	Glu	Ser	Gln	His	Val 220	Tyr	Tyr	Leu	Thr
Phe 225	Glu	Leu	Val	Leu	Met 230	Tyr	Сув	Asp	Val	Ile 235	Glu	Gly	Arg	Leu	Met 240
Thr	Glu	Thr	Ala	Met 245	Thr	Ile	Asp	Ala	Arg 250	Tyr	Ala	Glu	Leu	Leu 255	Gly
Arg	Val	Arg	Tyr 260	Met	Trp	Lys	Leu	11e 265	Asp	Gly	Phe	Phe	Pro 270	Ala	Leu
Gly	Asn	Pro 275	Thr	Tyr	Gln	Ile	Val 280	Ala	Met	Leu	Glu	Pro 285	Leu	Ser	Leu
Ala	Tyr 290	Leu	Gln	Leu	Arg	Asp 295	Ile	Thr	Val	Glu	Leu 300	Arg	Gly	Ala	Phe
Leu 305	Asn	His	Сув	Phe	Thr 310	Glu	Ile	His	Asp	Val 315	Leu	Asp	Gln	Asn	Gly 320
Phe	Ser	Asp	Glu	Gly 325	Thr	Tyr	His	Glu	Leu 330	Ile	Glu	Ala	Leu	Авр 335	Tyr
Ile	Phe	Ile	Thr 340	Asp	увр	Ile	His	Leu 345	Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
Phe	Arg	Ser 355	Phe	Gly	Ris	Pro	Arg 360	Leu	Glu	Ala	Val	Thr 365	Ala	Ala	Glu

- 155 -

Asn	. Val 370		L Lys	Туз	Met	375		Pro	Lye	Va]	11e 380		1 Туз	Glu	Thr
Leu 385		Lya	Gly	Hie	390		Phe	Сув	Gly	7 Ile 395		Ile	а Хвт	Gly	Tyr 400
Arg	Asp	Arg	His	Gly 405		Ser	Trp	Pro	Pro 410		Thr	Let	ı Pro	415	His
Ala	Ala	Asp	120		Arg	Asn	Ala	Gln 425		Ser	Gly	Glu	430		Thr
His	Glu	Gln 435		Val	Asp	Asn	Trp 440		Ser	Phe	Ala	Gly 445		Arg	Phe
Gly	Сув 450	Phe	Met	Pro	Leu	Ser 455	Leu	Asp	Ser	Asp	<b>Leu</b> 460		Met	Tyr	Leu
465					470					475					Tyr 480
				485	Arg				490					495	
			500		Phe			505					510		
		515			Val		520					525			
	530				Leu	535					540				
545					Thr 550					555					560
				565	Asn				570					575	
			580		His			585					590		
		595			Lys		600					605			
	610				Ser	615					620				
525					Gly 630 His					635					640
		-		- 4								- 7 -			

- 156 -

				645					650					655	
Ser	Ala	Phe	Ile 660	Thr	Thr	Двр	Leu	<b>L</b> ув 665	ГÀв	Tyr	Сув	Leu	Asn 670	Trp	Arg
Tyr	Glu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685	Ile	Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	Leu 700	Glu	Thr	Ser	Val
705	Tyr				710					715					720
	Leu			725					730				-	735	
	Gly		740					745					750		
	Tyr	755					760					765			
	Val 770					775					780				
785	Thr				790					795					800
	Авр			805					810					815	
	Leu		820					825					830		
	Lys	835					840					845			
	11e 850					855					860				
865	Ala				870					875					880
	Gly			885					890					895	
	Gln		900					905					910		
Thr	Arg	Авр 915	Val	Val	Ile		Leu 920	Leu	Thr	Asn	Asn	Asp 925	Leu	Leu	Ile

- 157 -

Arg	Met 930	Ala	Leu	Leu	Pro	Ala 935	Pro	Ile	Gly	Gly	Met 940		Tyr	Leu	Asn
Met 945	Ser	Arg	Leu	Phe	Val 950	Arg	Asn	Ile	Gly	Asp 955	Pro	Val	Thr	Ser	Ser 960
Ile	Ala	Asp	Leu	Lув 965	Arg	Met	Ile	Leu	Ala 970	Ser	Leu	Met	Pro	Glu 975	Glu
Thr	Leu	His	Gln 980	Val	Met	Thr	Gln	Gln 985	Pro	Gly	Asp	Ser	Ser 990	Phe	Leu
Asp	Trp	Ala 995	Ser	Asp	Pro	Tyr	Ser 100		Asn	Leu	Val	Сув 100		Gln	Ser
Ile	Thr 101		Leu	Leu	Lys	Asn 101		Thr	Ala	Arg	Phe 102		Leu	Ile	His
Ser 102		Asn	Pro	Met	Leu 103		Gly	Leu	Phe	His 103		Авр	Ser	Lys	Glu 1040
Glu	Asp	Glu	Arg	Leu 104!		Ala	Phe	Leu	Met 105	Asp 0	Arg	His	Ile	Ile 1055	
Pro	Arg	Ala	Ala 1060		Glu	Ile	Leu	Asp 106		Ser	Val	Thr	Gly 1070		Arg
Glu	Ser	Ile 1075		Gly	Met	Leu	Asp 1080		Thr	Lys	Gly	Leu 1085		Arg	Ala
	1090	)				1095	5			Val	1100	)	_		
Asn 1105		Авр	Tyr	Glu	Gln 1110		Arg	Ala	Gly	Met 1115		Leu	Leu	Thr	Gly 1120
Arg	Lys	Arg	Asn	Val 1125		Ile	Asp	Lys	Glu 1130	Ser )	Сув	Ser	Val	Gln 1135	
Ala	Arg	Ala	Leu 1140		Ser	His	Met	Trp 1145		Arg	Leu	Ala	Arg 1150		Arg
Pro	Ile	Tyr 1155		Leu	Glu	Val	Pro 1160		Val	Leu	Glu	Ser 1165		Arg	Gly
His	Leu 1170		Arg	Arg	His	Glu 1175		Сув	Val	Ile	Сув 1180		Сув	Gly	Ser
Val 1185	Aen	Tyr	Gly	Trp	Phe 1190		Val	Pro	Ser	Gly 1195		Gln	Leu	Asp	Авр 1200

- 158 -

- The Asp Lys Glu Thr Ser Ser Leu Arg Val Pro Tyr Ile Gly Ser Thr 1205 1210 1215
- Thr Asp Glu Arg Thr Asp Met Lys Leu Ala Phe Val Arg Ala Pro Ser 1220 1225 1230
- Arg Ser Leu Arg Ser Ala Val Arg Ile Ala Thr Val Tyr Ser Trp Ala 1235 1240 1245
- Tyr Gly Asp Asp Ser Ser Trp Asn Glu Ala Trp Leu Leu Ala Arg 1250 1255 1260
- Gln Arg Ala Asn Val Ser Leu Glu Glu Leu Arg Val Ile Thr Pro Ile 1265 1270 1275 1280
- Ser Thr Ser Thr Asn Leu Ala His Arg Leu Arg Asp Arg Ser Thr Gln
  1285 1290 1295
- Val Lys Tyr Ser Gly Thr Ser Leu Val Arg Val Ala Arg Tyr Thr Thr 1300 1305 1310
- Ile Ser Asn Asp Asn Leu Ser Phe Val Ile Ser Asp Lys Lys Val Asp 1315 1320 1325
- Thr Asn Phe Ile Tyr Gln Gln Gly Met Leu Leu Gly Leu Gly Val Leu 1330 1335 1340
- Glu Thr Leu Phe Arg Leu Glu Lys Asp Thr Gly Ser Ser Asn Thr Val 1345 1350 1355 1360
- Leu His Leu His Val Glu Thr Asp Cys Cys Val Ile Pro Met Ile Asp 1365 1370 1375
- His Pro Arg Ile Pro Ser Ser Arg Lys Leu Glu Leu Arg Ala Glu Leu 1380 1385 1390
- Cys Thr Asn Pro Leu Ile Tyr Asp Asn Ala Pro Leu Ile Asp Arg Asp 1395 1400 1405
- Ala Thr Arg Leu Tyr Thr Gln Ser His Arg Arg His Leu Val Glu Phe 1410 1415 1420
- Val Thr Trp Ser Thr Pro Gln Leu Tyr His Ile Leu Ala Lys Ser Thr 1425 1430 1435 1440
- Ala Leu Ser Met Ile Asp Leu Val Thr Lys Phe Glu Lys Asp His Met 1445 1450 1455
- Asn Glu Ile Ser Ala Leu Ile Gly Asp Asp Asp Ile Asn Ser Phe Ile 1460 1465 1470
- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly

- 159 -

		1475	5				1480	)				1485	5		
Gln	Cys 1490		Ala	Ile	Asn	Trp 1495		Phe	Asp	Val	His 1500	-	His	Arg	Pro
Ser 1505	-	Lys	Tyr	Gln	Met 1510	_	Glu	Leu	Leu	Ser 1515		Phe	Leu	Ser	Arg 1520
Met	Ser	Lys	Gly	Val 1525		Lys	Val	Leu	Val 1530		Ala	Leu	Ser	His 1535	
Lys	Ile	Tyr	Lys 1540	Lys )	Phe	Trp	His	Cys. 1545	-	Ile	Ile	Glu	Pro 1550		His
Gly	Pro	Ser 1555		Asp	Ala	Gln	Asn 1560		His	Thr	Thr	Val 1565	•	Asn	Met
Val	Tyr 1570		Сув	Tyr	Met	Thr 1575	_	Leu	Asp	Leu	Leu 1580		Asn	Glu	Glu
<b>Le</b> u 1585		Glu	Phe	Thr	Phe 1590		Leu	Cys	Glu	Ser 1595	_	Glu	Asp	Val	Val 1600
Pro	Asp	Arg	Phe	Asp 1605		Ile	Gln	Ala	Lys 1610		Leu	Сув	Val	Leu 1615	
Asp	Leu	Tyr	Сув 1620	Gln	Pro	Gly	Thr	Сув 1625		Pro	Ile	Arg	Gly 1630		Arg
Pro	Val	Glu 1635	-	Сув	Ala	Val	Leu 1640		Asp	His	Ile	Lys 1645		Glu	Ala
Arg	Leu 1650		Pro	Ala	Gly	Ser 1655		Trp	Asn	Ile	Asn 1660		Ile	Ile	Val
Авр 1665		Tyr	Ser	Сув	Ser 1670		Thr	Tyr	Leu	Arg 1675	_	Gly	Ser	Ile	Lys 1680
Gln	Ile	Arg	Leu	Arg 1685		ĄsĄ	Pro	Gly	Phe 1690		Phe	qaA	Ala	Leu 1695	
Glu	Val	Asn	Val 1700	Ser	Gln	Pro	Lys	Val 1705	_	Ser	Asn	Asn	Ile 1710		Asn
Met	Ser	Ile 1715	-	увр	Phe	_	Pro 1720		His	Asp	Asp	Val 1725		Lys	Leu
Leu	Lys 1730	-	Ile	Asn	Thr	Ser 1735	-	His	Asn	Leu	Pro 1740		Ser	Gly	Gly
Ser 1745		Ala	Asn	Tyr	Glu 1750		His	Ala	Phe	Arg 1755	_	Ile	Gly	Leu	Asn 1760

- 160 -

- Ser Ser Ala Cys Tyr Lys Ala Val Glu Ile Ser Thr Leu Ile Arg Arg 1765 1770 1775
- Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly 1780 1785 1790
- Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe 1795 1800 1805
- Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu 1810 1815 1820
- Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val 1825 1830 1835 1840
- Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp 1845 1850 1855
- Val Gly Ser Ile Asp Cys Phe Asn Phe Ile Val Ser Asn Ile Pro Thr 1860 1865 1870
- Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asn Lys 1875 1880 1885
- Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala 1890 1895 1900
- Leu Leu Cly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro 1905 1910 1915 1920
- Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser His
  1925 1930 1935
- Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser 1940 1945 1950
- Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965
- Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1975 1980
- Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000
- Ile Gln Ala Ile Val Gly Gly Ala Val Ser Arg Gly Asp Ile Asn Pro
  2005 2010 2015
- Ile Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Ser Cys Gly
  2020 2025 2030

- 161 -

	Leu	Ala	Ile 2035		Gly	Pro	Lys	Leu 2040		Lys	Glu	Leu	11e 2045		His	Asp	
	Val	Ala 2050		Gly	Gln	Asp	Gly 205		Leu	Asn	Ser	Ile 2060		Ile	Leu	Tyr	
	Arg 2065		Leu	Ala	Arg	Phe 2070	Lys	Авр	Asn	Gln	Arg 2075		Gln	Gln	Gly	Met 2080	
	Phe	His	Ala	Tyr	Pro 2085		Leu	Val	Ser	Ser 2090		Gln	λrg	Glu	Leu 2095		
	Ser	Arg	Ile	Thr 210	_	Lys	Phe	Trp	Gly 210		Ile	Leu	Leu	Tyr 2110		Gly	
	Asn	Arg	Lys 211		Ile	Asn	Arg	Phe 212		Gln	Asn	Leu	Lув 212		Gly	Tyr	
	Leu	Ile 213		Авр	Leu	His	Gln 213		Ile	Phe	Val	Lys 214		Leu	Ser	Lys	
	Ser 2145		Lys	Gln	Ile	Ile 215	Met 0	Thr	Gly	Gly	Leu 215	-	Arg	Glu	Trp	Val 2160	
	Phe	ГÀВ	Val	Thr	Val 216		Glu	Thr	Lys	Glu 217(		Tyr	Lys	Leu	Val 2175		
	Tyr	Ser	Ala	Leu 218		Lys	Asp										
(2) 1	NFO	RMAT:	ION 1	FOR S	SEQ :	ID N	0:9:										
	(i)	(A) (B) (C)	LEI TY:	ngth Pe: 1 Randi	: 150 nucle EDNE:	894 1 Bic :	STIC: base acid sing: ar	pai	rs								
,	(ii)	MOL	BCOL:	E TY	PE: 1	RNA	(gen	omic	)								
	(xi)	SEQ	denc:	e de:	SCRI	PTIO	N: S	EQ II	ои с	:9:							
ACCA	ACA	AA G	TTGG	CTAA	G GA	TAGT	TCAA	TCA	ATGA:	rca :	CTT	CTAG'	rg C	ACTT	AGGA:	r	60
TCAA	SATC	CT A	TAT	CAGG	g ac	AAGA	GCAG	GAT:	ragg(	GAT 1	ATCC	gaga'	rg go	CAC	CTT	r	120
TAAGO	BAGC	TT A	GCAT:	rgtt	CAŅ	aaga:	AACA	AGG	ACAA	ACC 2	ACCC	ATTA(	CA TO	CAGG	ATCC	3	180
GTGG	AGCC	AT C	AGAG	GAAT	C AA	ACAC	ATTA	TTA:	ragt:	ACC 2	AATC	CCTG	GA G	ATTC	CTCAI	A	240

- 162 -

TTACCACTCG A	TCCAGACTT	CTGGACCGG	г тестелест	r aammaaaaa	C CCGGATGTGA	
				•		300
GCGGGCCCAA A	CTAACAGGG	GCACTAATA	GTATATTAT(	CTTATTTGT	G GAGTCTCCAG	360
GTCAATTGAT 1	CAGAGGATC	ACCGATGAC	CTGACGTTAG	CATAAGGCT	TTAGAGGTTG	420
TCCAGAGTGA C	Cagtcacaa	TCTGGCCTT	CCTTCGCATC	AAGAGGTAC	AACATGGAGG	480
ATGAGGCGGA C	CAATACTTT	TCACATGATO	ATCCAATTAG	TAGTGATCA	TCCAGGTTCG	540
GATGGTTCGA G	AACAAGGAA	ATCTCAGATA	TTGAAGTGCA	AGACCCTGAG	GGATTCAACA	600
TGATTCTGGG T	ACCATCCTA	GCCCAAATTT	GGGTCTTGCT	CGCAAAGGC	GTTACGGCCC	660
CAGACACGGC A	GCTGATTCG	GAGCTAAGAA	GGTGGATAAA	GTACACCCA	CAAAGAAGGG	720
TGGTTGGTGA A	TTTAGATTG	GAGAGAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC C	TTACGCCGA	TTCATGGTCG	CTCTAATCCT	GGATATCAAG	AGAACACCCG	840
GAAACAAACC C	AGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATO	GTAGAGGCAG	900
GATTAGCCAG T	TTTATCCTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA A	TTTGCT <b>G</b> GT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAC	CTTTACCAGC	1020
AAATGGGGGA AI	ACTGCACCC	TACATGGTAA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATC AT	PACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG AC	GGTTTGAAC	TTTGGCCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT G	STAAGGAGG	TCAGCTGGAA	AGGTCAGTTC	CACATTGGCA	TCTGAACTCG	1260
GTATCACTGC CO	AGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCATACT	ACTGAGGACA	1320
AGATCAGTAG AG	CGGTTGGA	CCCAGACAAG	CCCAAGTATC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA GC	TACCGAGA	TTGGGGGGCA	AGGAAGATAG	GAGGGTCAAA	CAGAGTCGAG	1440
GAGAAGCCAG GG	AGAGCTAC .	AGAGAAACCG	GGCCCAGCAG	AGCAAGTGAT	GCGAGAGCTG	1500
CCCATCTTCC AA	CCGGCACA	CCCCTAGACA	TTGACACTGC	AACGGAGTCC	AGCCAAGATC	1560
CGCAGGACAG TC	GAAGGTCA	GCTGACGCCC	TGCTTAGGCT	GCAAGCCATG	GCAGGAATCT	1620
CGGAAGAACA AG	GCTCAGAC	ACGGACACCC	CTATAGTGTA	CAATGACAGA	AATCTTCTAG	1680
ACTAGGTGCG AG	AGGCCGAG	GGCCAGAACA	ACATCCGCCT	ACCCTCCATC	ATTGTTATAA	1740
AAAACTTAGG AA	CCAGGTCC 1	ACACAGCCGC	CAGCCCATCA	ACCATCCACT	CCCACGATTG	1800

- 163 -

GAGCCAATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	ATCGAGGAAG	CTATGGCAGC	ATGGTCAGAA	1920
ATATCAGACA	ACCCAGGACA	GGAGCGAGCC	ACCTGCAGGG	AAGAGAAGGC	AGGCAGTTCG	1980
GGTCTCAGCA	AACCATGCCT	CTCAGCAATT	GGATCAACTG	AAGGCGGTGC	ACCTCGCATC	2040
CGCGGTCAGG	GACCTGGAGA	GAGCGATGAC	GACGCTGAAA	CTTTGGGAAT	CCCCCAAGA	2100
AATCTCCAGG	CATCAAGCAC	TGGGTTACAG	TGTTATTACG	TTTATGATCA	CAGCGGTGAA	2160
GCGGTTAAGG	GAATCCAAGA	TGCTGACTCT	ATCATGGTTC	AATCAGGCCT	TGATGGTGAT	2220
AGCACCCTCT	CAGGAGGAGA	CAATGAATCT	GAAAACAGCG	ATGTGGATAT	TGGCGAACCT	2280
GATACCGAGG	GATATGCTAT	CACTGACCGG	GGATCTGCTC	CCATCTCTAT	GGGGTTCAGG	2340
GCTTCTGATG	TTGAAACTGC	AGAAGGAGGG	GAGATCCACG	AGCTCCTGAG	ACTCCAATCC	2400
AGAGGCAACA	ACTTTCCGAA	GCTTGGGAAA	ACTCTCAATG	TTCCTCCGCC	CCCGGACCCC	2460
GGTAGGGCCA	GCACTTCCGG	GACACCCATT	AAAAAGGGCA	CAGACGCGAG	ATTAGCCTCA	2520
TTTGGAACGG	AGATCGCGTC	TTTATTGACA	GGTGGTGCAA	CCCAATGTGC	TCGAAAGTCA	2580
CCCTCGGAAC	CATCAGGGCC	AGGTGCACCT	GCGGGGAATG	TCCCCGAGTG	TGTGAGCAAT	2640
GCCGCACTGA	TACAGGAGTG	GACACCCGAA	TCTGGTACCA	CAATCTCCCC	GAGATCCCAG	2700
AATAATGAAG	AAGGGGGAGA	CTATTATGAT	GATGAGCTGT	TCTCTGATGT	CCAAGATATT	2760
AAAACAGCCT	TGGCCAAAAT	ACACGAGGAT	AATCAGAAGA	TAATCTCCAA	GCTAGAATCA	2820
CTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAGAAGC	AGATCAACAG	GCAAAATATC	2880
AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TCGCCATTCC	TGGACTTGGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTC	GAAATCAATC	CCGACTTGAA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AACCCGTTGC	CAGCCGACAA	3060
CTCCAAGGAA	TGACAAATGG	ACGGACCAGT	TCCAGAGGAC	AGCTGCTGAA	GGAATTTCAG	3120
CTAAAGCCGA	TCGGGAAAAA	GATGAGCTCA	GCCGTCGGGT	TTGTTCCTGA	CACCGGCCCT	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGA	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ATGATCTTGC	CAAGTTCCAC	3300
CAGATGCTGA	TGAAGATAAT	AATGAAGTAG	CTACAGCTCA	ACTTACCTGC	CAACCCCATG	3360

- 164 -

CCAGTCGAC	C CAACTAGTA	C AACCTAAAT	C CATTATAAA	A AACTTAGGA	G CAAAGTGATT	3420
GCCTCCCAA	G TTCCACAAT	g acagagacc	T ACGACTTCG	A CAAGTCGGC	A TGGGACATCA	3480
AAGGGTCGA	T CGCTCCGAT	A CAACCCACC	A CCTACAGTG	A TGGCAGGCT	GTGCCCCAGG	3540
TCAGAGTCA	T AGATCCTGG:	T CTAGGCGAC	A GGAAGGATG	A ATGCTTTATO	F TACATGTTTC	3600
TGCTGGGGG	r tgttgågga	C AGCGATTCC	C TAGGGCCTCC	AATCGGGCG!	GCATTTGGGT	3660
CCCTGCCCT	r aggtgttgg(	C AGATCCACA(	G CAAAGCCCG	AAAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTG	A CATAGTTGTT	AGACGTACAC	G CAGGGCTCAR	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCC	ACTAACTCT	CTCACACCT	GGAGAAAGGI	CCTAACAACA	GGGAGTGTCT	3840
TCAACGCAAI	CCAAGTGTGC	: AATGCGGTTA	ATCTGATACO	GCTCGATACC	CCGCAGAGGT	3900
TCCGTGTTG	TTATATGAGO	ATCACCCGTC	TTTCGGATAA	CGGGTATTAC	ACCGTTCCTA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	GGCGATAGGC	CCTGGGAAGA	TCATCGACAA	TACAGAGCAA	CTTCCTGAGG	4080
CAACATTTAI	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	GAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCTTATGT	TACCCGCTGA	TGGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380
AAGAATTCCG	CATTTACGAC	GACGTGATCA	TAAATGATGA	CCAAGGACTA	TTCAAAGTTC	4440
TGTAGACCGT	AGTGCCCAGC	AATGCCCGAA	AACGACCCCC	CTCACAATGA	CAGCCAGAAG	4500
GCCCGGACAA	AAAAGCCCCC	TCCGAAAGAC	TCCACGGACC	AAGCGAGAGG	CCAGCCAGCA	4560
GCCGACGGCA	AGCGCGAACA	CCAGGCGGCC	CCAGCACAGA	ACAGCCCCGA	CACAAGGCCA	4620
CCACCAGCCA	CCCCAATCTG	CATCCTCCTC	GTGGGACCCC	CGAGGACCAA	CCCCCAAGGC	4680
TGCCCCCGAT	CCAAACCACC	AACCGCATCC	CCACCACCCC	CGGGAAAGAA	ACCCCCAGCA	4740
ATTGGAAGGC	CCCTCCCCCT	CTTCCTCAAC	ACAAGAACTC	CACAACCGAA	CCGCACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGATCCTCTC	TCCCCGGCAA	4860
ACTAAACAAA	ACTTAGGGCC	AAGGAACATA	CACACCCAAC	AGAACCCAGA	CCCCGGCCCA	4920

- 165 -

CGGCGCCGCG CCCCCAACCC CCGACAACCA GAGGGAGCCC CCAACCAA	4980
CCCCGGTGCC CACAGGCAGG GACACCAACC CCCGAACAGA CCCAGCACCC AACCATCGAC	5040
AATCCAAGAC GGGGGGGCCC CCCCAAAAAA AGGCCCCCAG GGGCCGACAG CCAGCACCGC	5100
GAGGAAGCCC ACCCACCCCA CACACGACCA CGGCAACCAA ACCAGAACCC AGACCACCCT	5160
GGGCCACCAG CTCCCAGACT CGGCCATCAC CCCGCAGAAA GGAAAGGCCA CAACCCGCGC	5220
ACCCCAGCCC CGATCCGGCG GGGAGCCACC CAACCCGAAC CAGCACCCAA GAGCGATCCC	5280
CGAAGGACCC CCGAACCGCA AAGGACACCA GTATCCCACA GCCTCTCCAA GTCCCCCGGT	5340
CTCCTCCTCT TCTCGAAGGG ACCAAAAGAT CAATCCACCA CACCCGACGA CACTCAACTC	5400
CCCACCCCTA AAGGAGACAC CGGGAATCCC AGAATCAAGA CTCATCCAAT GTCCATCATG	5460
GGTCTCAAGG TGAACGTCTC TGCCATATTC ATGGCAGTAC TGTTAACTCT CCAAACACCC	5520
ACCGGTCAAA TCCATTGGGG CAATCTCTCT AAGATAGGGG TGGTAGGAAT AGGAAGTGCA	5580
AGCTACAAAG TTATGACTCG TTCCAGCCAT CAATCATTAG TCATAAAATT AATGCCCAAT	5640
ATAACTCTCC TCAATAACTG CACGAGGGTA GAGATTGCAG AATACAGGAG ACTACTGAGA	5700
ACAGTTTTGG AACCAATTAG AGATGCACTT AATGCAATGA CCCAGAATAT AAGACCGGTT	5760
CAGAGTGTAG CTTCAAGTAG GAGACACAAG AGATTTGCGG GAGTAGTCCT GGCAGGTGCG	5820
GCCCTAGGCG TTGCCACAGC TGCTCAGATA ACAGCCGGCA TTGCACTTCA CCAGTCCATG	5880
CTGAACTCTC AAGCCATCGA CAATCTGAGA GCGAGCCTGG AAACTACTAA TCAGGCAATT	5940
GAGGCAATCA GACAAGCAGG GCAGGAGATG ATATTGGCTG TTCAGGGTGT CCAAGACTAC	6000
ATCAATAATG AGCTGATACC GTCTATGAAC CAACTATCTT GTGATTTAAT CGGCCAGAAG	6060
CTCGGGCTCA AATTGCTCAG ATACTATACA GAAATCCTGT CATTATTTGG CCCCAGTTTA	6120
CGGGACCCCA TATCTGCGGA GATATCTATC CAGGCTTTGA GCTATGCGCT TGGAGGAGAC	6180
ATCAATAAGG TGTTAGAAAA GCTCGGATAC AGTGGAGGTG ATTTACTGGG CATCTTAGAG	6240
AGCGGAGGAA TAAAGGCCCG GATAACTCAC GTCGACACAG AGTCCTACTT CATTGTCCTC	6300
AGTATAGCCT ATCCGACGCT GTCCGAGATT AAGGGGGTGA TTGTCCACCG GCTAGAGGGG GTCTCGTACA ACATAGGCTC TCAAGAGTGG TATACCACTG TGCCCAAGTA TGTTGCAACC	6360
CAAGGGTACC TTATCTCGAA TTTTGATGAG TCATCGTGTA CTTTCATGCC AGAGGGGACT	6420
THE TOTAL CITICAL OCC. AGAGGGGACT	6480

- 166 -

GTGTGCAGCC	: AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGGG	6540
TACACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GATCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGACGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATCATTGG	AGAGGTTGGA	CGTAGGGACA	6840
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	AATTGTTGGA	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTATCG	AGCACTAGCA	TAGTCTACAT	CCTGATTGCA	6960
GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACGGGA	7080
ACATCAAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCACCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	AGTAGGATAG	TCATTAACAG	AGAACATCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTTG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
CATCTCTGAC	AAGATTAAAT	TCCTTAATCC	GGATAGGGAG	TACGACTTCA	GAGATCTCAC	7680
TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
GGCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
ATTCTCAAAC	ATGTCGCTGT	CCCTGTTAGA	CTTGTATTTA	GGTCGAGGTT	ACAATGTGTC	7920
					TGGAAAAGCC	
TAATCTGAGC	AGCAAAAGGT	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040

- 167 -

AGGTGTTATO	CAGAAATCCGG	GTTTGGGGG	C TCCGGTGTT	C CATATGACA	A ACTATCTTGA	8100
GCAACCAGT	AGTAATGAT	CTCAGCAACTO	TATGGTGGC	TTGGGGGAG	TCAAACTCGC	8160
AGCCCTTTGT	CACGGGGAAG	ATTCTATCA	C AATTCCCTAT	CAGGGATCA	GGAAAGGTGT	8220
CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCT	GAAATCCCC	ACCGACATGO	AATCCTGGGT	8280
CCCCTTATCA	ACGGATGATO	CAGTGATAGA	CAGGCTTTAC	CTCTCATCTC	ACAGAGGTGT	8340
TATCGCTGAC	AATCAAGCAA	AATGGGCTGT	CCCGACAACA	CGAACAGATG	ACAAGTTGCG	8400
AATGGAGACA	TGCTTCCAAC	: AGGCGTGTAA	GGGTAAAATC	CAAGCACTCI	GCGAGAATCC	8460
CGAGTGGGCA	CCATTGAAGG	ATAACAGGAT	TCCTTCATAC	GGGGTCTTGT	CTGTTGATCT	8520
GAGTCTGACA	GTTGAGCTTA	AAATCAAAAT	TGCTTCGGGA	TTCGGGCCAT	TGATCACACA	8580
CGGTTCAGGG	ATGGACCTAT	ACAAATCCAA	CCACAACAAT	GTGTATTGGC	TGACTATCCC	8640
GCCAATGAAG	AACCTAGCCT	TAGGTGTAAT	CAACACATTG	GAGTGGATAC	CGAGATTCAA	8700
GGTTAGTCCC	TACCTCTTCA	CTGTCCCAAT	TAAGGAAGCA	GGCGAAGACT	GCCATGCCCC	8760
AACATACCTA	CCTGCGGAGG	TGGATGGTGA	TGTCAAACTC	AGTTCCAATC	TGGTGATTCT	8820
ACCTGGTCAA	GATCTCCAAT	ATGTTTTGGC	AACCTACGAT	ACTTCCAGGG	TTGAACATGC	8880
TGTGGTTTAT	TACGTTTACA	GCCCAAGCCG	CTCATTTTCT	TACTTTTATC	CTTTTAGGTT	8940
GCCTATAAAG	GGGGTCCCCA	TCGAATTACA	AGTGGAATGC	TTCACATGGG	ACCAAAAACT	9000
CTGGTGCCGT	CACTTCTGTG	TGCTTGCGGA	CTCAGAATCT	GGTGGACATA	TCACTCACTC	9060
TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACCCGG	GAAGATGGAA	CCAATCGCAG	9120
ATAGGGCTGC	TAGTGAACCA	ATCACATGAT	GTCACCCAGA	CATCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATT	AAGAAAAACG	TAGGGTCCAA	GTGGTTCCCC	GTTATGGACT	9240
CGCTATCTGT	CAACCAGATC	TTATACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
ATAAGATAGT	AGCCATCCTG	GAGTATGCTC	GAGTCCCTCA	CGCTTACAGC	CTGGAGGACC	9360
CTACACTGTG	TCAGAACATC	AAGCACCGCC	TAAAAAACGG	ATTTTCCAAC	CAAATGATTA	9420
TAAACAATGT	GGAAGTTGGG	AATGTCATCA	AGTCCAAGCT	TAGGAGTTAT	CCGGCCCACT	9480
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTTAA	CATAGAAGAC	aaagagtcaa	9540
CGAGGAAGAT	CCGTGAACTC	CTCAAAAAGG	GGAATTCGCT	GTACTCCAAA	GTCAGTGATA	9600

- 168 -

AGGTTTTCC	A ATGCTTAAGG	GACACTAACT	CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCA	a ggagaaagtt	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAGTGGTTTG	9720
AGCCCTTTC	T GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTT	G CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCT	C TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TGACATTTG	A ACTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGA	C TATTGATGCT	AGGTATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AACTGATAG	A TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCCATGC	10080
TGGAGCCTC	T TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTA	A CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
ATGAAGGTA	C TTATCATGAG	TTAATTGAAG	CTCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGA	C AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGG	C TGCTGAAAAT	GTTAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGA	T GAAAGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAG	G CAGTTGGCCA	CCGCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAAG	C TTCAGGTGAA	GGGTTAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAAATCTT	10560
TTGCTGGAG	T GAAATTTGGC	TGCTTTATGC	CTCTTAGCCT	GGATAGTGAT	CTGACAATGT	10620
ACCTAAAGG	A CAAGGCACTT	GCTGCTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680
AGTTCCTGC	G TTACGACCCT	CCCAAGGGAA	CCGGGTCACG	GAGGCTTGTA	GATGTTTTCC	10740
TTAATGATT	C GAGCTTTGAC	CCATATGATG	TGATAATGTA	TGTTGTAAGT	GGAGCTTACC	10800
TCCATGACC	C TGAGTTCAAC	CTGTCTTACA	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTTT	T TGCTAAAATG	ACTTACAAAA	TGAGGGCATG	CCAAGTGATT	GCTGAAAATC	10920
TAATCTCAA	A CGGGATTGGC	AAATATTTTA	AGGACAATGG	GATGGCCAAG	GATGAGCACG	10980
ATTTGACTA	A GGCACTCCAC	ACTCTAGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGGG	G GGGGCCAGTC	TTAAAAACCT	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100
GGAACGTGA	G AGCAGCAAAA	GGGTTTATAG	GGTTCCCTCA	AGTAATTCGG	CAGGACCAAG	11160

- 169 -

ACACTGATCA	TCCGGAGAAT	ATGGAAGCTT	ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAGAA	GTACTGCCTT	aattggagat	ATGAGACCAT	CAGCTTGTTT	GCACAGAGGC	11280
TAAATGAGAT	TTACGGATTG	CCCTCATTTT	TCCAGTGGCT	GCATAAGAGG	CTTGAGACCT	11340
CTGTCCTGTA	TGTAAGTGAC	CCTCATTGCC	CCCCGACCT	TGACGCCCAT	ATCCCGTTAT	11400
ATAAAGTCCC	CAATGATCAA	ATCTTCATTA	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTATCTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGGAT	TGCTTCGTTA	GTGCAAGGGG	ACAATCAGAC	CATAGCCGTA	ACAAAAAGGG	11580
TACCCAGCAC	ATGGCCCTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ATATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAA	ATTCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACAACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGGG	ATGAATTATC	12060
TGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCCTCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180
CACAACAACO	GGGGGACTCT	TCATTCCTAG	CTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTGTATGTGT	CCAGAGCATO	: ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTGA	12300
TCCATAGTCC	: AAACCCAATG	TTAAAAGGAT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGGGACTGGC	GGCATTCCTC	ATGGACAGG	ATATTATAGT	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATC	A TAGTGTCAC	GGGGCAAGA	3 AGTCTATTGC	AGGCATGCTG	GATACCACAA	12480
AAGGCTTGAT	TCGAGCCAG	ATGAGGAAG	G GGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATT	A TGACTATGAI	A CAATTCAGA	E CAGGGATGGT	GCTATTGACA	GGAAGAAAGC	12600
GAAATGTCC	r cattgacaai	A GAGTCATGT	r cagtgcagct	GGCGAGAGCT	CTAAGAAGCC	12660
ATATGTGGG	GAGGCTAGC	T CGAGGACGG	CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720

- 170 -

PCT/US97/16718

TAGAATCTAT GCGAGGCCAC CTTATTCGGC GTCATGAGAC ATGTGTCATC TGCGAGTGTG	
GATCAGTCAA CTACGGATGG TTTTTTGTCC CCTCGGGTTG CCAACTGGAT GATATTGACA	12840
AGGAAACATC ATCCTTGAGA GTCCCATATA TTGGTTCTAC CACTGATGAG AGAACAGACA	12900
TGAAGCTTGC CTTCGTAAGA GCCCCAAGTC GATCCTTGCG ATCTGCTGTT AGAATAGCAA	12960
CAGTGTACTC ATGGGCTTAC GGTGATGATG ATAGCTCTTG GAACGAAGCC TGGTTGTTGG	13020
CTAGGCAAAG GGCCAATGTG AGCCTGGAGG AGCTAAGGGT GATCACTCCC ATCTCAACTT	13080
CGACTAATTT AGCGCATAGG TTGAGGGATC GTAGCACTCA AGTGAAATAC TCAGGTACAT	13140
CCCTTGTCCG AGTGGCGAGG TATACCACAA TCTCCAACGA CAATCTCTCA TTTGTCATAT	13200
CAGATAAGAA GGTTGATACT AACTTTATAT ACCAACAAGG AATGCTTCTA GGGTTGGGTG	13260
TTTTAGAAAC ATTGTTTCGA CTCGAGAAAG ATACCGGATC ATCTAACACG GTATTACATC	13320
TTCACGTCGA AACAGATTGT TGCGTGATCC CGATGATAGA TCATCCCAGG ATACCCAGCT	13380
CCCGCAAGCT AGAGCTGAGG GCAGAGCTAT GTACCAACCC ATTGATATAT GATAATGCAC	13440
CTTTAATTGA CAGAGATGCA ACAAGGCTAT ACACCCAGAG CCATAGGAGG CACCTTGTGG	13500
AATTTGTTAC ATGGTCCACA CCCCAACTAT ATCACATTTT AGCTAAGTCC ACAGCACTAT	13560
CTATGATTGA CCTGGTAACA AAATTTGAGA AGGACCATAT GAATGAAATT TCAGCTCTCA	13620
TAGGGGATGA CGATATCAAT AGTTTCATAA CTGAGTTTCT GCTCATAGAG CCAAGATTAT	13680
TCACTATCTA CTTGGGCCAG TGTGCGGCCA TCAATTGGGC ATTTGATGTA CATTATCATA	13740
GACCATCAGG GAAATATCAG ATGGGTGAGC TGTTGTCATC GTTCCTTTCT AGAATGAGCA	13800
AAGGAGTGTT TAAGGTGCTT GTCAATGCTC TAAGCCACCC AAAGATCTAC AAGAAATTCT	13860
GGCATTGTGG TATTATAGAG CCTATCCATG GTCCTTCACT TGATGCTCAA AACTTGCACA	13920
CAACTGTGTG CAACATGGTT TACACATGCT ATATGACCTA CCTCGACCTG TTGTTGAATG	13980
AAGAGTTAGA AGAGTTCACA TTTCTCTTGT GTGAAAGCGA CGAGGATGTA GTACCGGACA	14040
GATTCGACAA CATCCAGGCA AAACACTTAT GTGTTCTGGC AGATTTGTAC TGTCAACCAG	14100
GGACCTGCCC ACCAATTCGA GGTCTAAGAC CGGTAGAGAA ATGTGCAGTT CTAACCGACC	14160
ATATCAAGGC AGAGGCTATG TTATCTCCAG CAGGATCTTC GTGGAACATA AATCCAATTA	14220
TTGTAGACCA TTACTCATGC TCTCTGACTT ATCTCCGGCG AGGATCGATC AAACAGATAA	14280

- 171 -

GA:	TTGAGAGT	TGATCCAGGA	TTCATTTTCG	ACGCCCTCGC	TGAGGTAAAT	GTCAGTCAGC	14340
CAI	AAGATCGG	CAGCAACAAC	ATCTCAAATA	TGAGCATCAA	GGCTTTCAGA	CCCCCACACG	14400
ATO	GATGTTGC	AAAATTGCTC	AAAGATATCA	ACACAAGCAA	GCACAATCTT	CCCATTTCAG	14460
GG	GGCAATCT	CGCCAATTAT	GAAATCCATG	CTTTCCGCAG	AATCGGGTTG	AACTCATCTG,	14520
CT:	igctacaa	AGCTGTTGAG	ATATCAACAT	TAATTAGGAG	ATGCCTTGAG	CCAGGGGAGG	14580
ACC	GCTTGTT	CTTGGGTGAG	GGATCGGGTT	CTATGTŢGAT	CACTTATAAG	GAGATACTTA	14640
AA	CTAAACAA	GTGCTTCTAT	AATAGTGGGG	TTTCCGCCAA	TTCTAGATCT	GGTCAAAGGG	14700
AA!	<b>TTAGCACC</b>	CTATCCCTCC	GAAGTTGGCC	TTGTCGAACA	CAGAATGGGA	GTAGGTAATA	14760
TT	GTCAAAGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	GTAGATTGCT	14820
TC	AATTTCAT	AGTTAGTAAT	ATCCCTACCT	CTAGTGTGGG	GTTTATCCAT	TCAGATATAG	14880
AG	ACCTTGCC	TGACAAAGAT	ACTATAGAGA	AGCTAGAGGA	ATTGGCAGCC	ATCTTATCGA	14940
TG	GCTCTGCT	CCTGGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
GG	GATTTTGT	TCAGGGATTT	ATAAGTTATG	TAGGGTCTCA	TTATAGAGAA	GTGAACCTTG	15060
TA	TACCCTAG	ATACAGCAAC	TTCATCTCTA	CTGAATCTTA	TTTGGTTATG	ACAGATCTCA	15120
AG	GCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	GATAATTGAA	TCATCTGTGA	15180
GG	ACTTCACC	TGGACTTATA	GGTCACATCC	TATCCATTAA	GCAACTAAGC	TGCATACAAG	15240
CA	attgtggg	AGACGCAGTT	AGTAGAGGTG	ATATCAATCC	TACTCTGAAA	AAACTTACAC	15300
CT	ATAGAGCA	GGTGCTGATC	AATTGCGGGT	TGGCAATTAA	CGGACCTAAG	CTGTGCAAAG	15360
AA:	TTGATCCA	CCATGATGTT	GCCTCAGGGC	AAGATGGATT	GCTTAATTCT	ATACTCATCC	15420
TC	TACAGGGA	GTTGGCAAGA	TTCAAAGACA	ACCAAAGAAG	TCAACAAGGG	ATGTTCCACG	15480
CT:	TACCCCGT	ATTGGTAAGT	AGCAGGCAAC	GAGAACTTAT	ATCTAGGATC	ACCCGCAAAT	15540
TC	TGGGGGCA	CATTCTTCTT	TACTCCGGGA	ACAGAAAGTT	GATAAATAAG	TTTATCCAGA	15600
AT	CTCAAGTC	CGGCTATCTG	ATACTAGACT	TACACCAGAA	TATCTTCGTT	AAGAATCTAT	15660
CC	aagtcaga	GAAACAGATT	ATTATGACGG	GGGGTTTGAA	ACGTGAGTGG	GTTTTTAAGG	15720
TA	ACAGTCAA	GGAGACCAAA	GAATGGTATA	AGTTAGTCGG	ATACAGTGCC	CTGATTAAGG	15780
AC'	TAATTGGT	TGAACTCCGG	AACCCTAATC	CTGCCCTAGG	TGGTTAGGCA	TTATTTGCAA	15840

- 172 -

# TATATTAAAG AAAACTITGA AAATACGAAG TITCTATTCC CAGCTITGTC TGGT

15894

- (2) INFORMATION FOR SEQ ID NO:10:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2183 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
  - Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15
  - Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala 20 25 30
  - Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45
  - Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 55 60
  - Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80
  - Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn 85 90 95
  - Ile Glu Asp Lys Glu Ser Thr Arg Lys Ile Arg Glu Leu Leu Lys Lys
    100 105 110
  - Gly Asn Ser Leu Tyr Ser Lys Val Ser Asp Lys Val Phe Gln Cys Leu 115 120 125
  - Arg Asp Thr Asn Ser Arg Leu Gly Leu Gly Ser Glu Leu Arg Glu Asp 130 135 140
  - Ile Lys Glu Lys Val Ile Asn Leu Gly Val Tyr Met His Ser Ser Gln
    145 150 155 160
  - Trp Phe Glu Pro Phe Leu Phe Trp Phe Thr Val Lys Thr Glu Met Arg
  - Ser Val Ile Lys Ser Gln Thr His Thr Cys His Arg Arg Arg His Thr

- 173 -

			180					185					190	)	
Pro	Val	Phe 195		Thr	Gly	Ser	Ser 200		Glu	Leu	Leu	11e 205		Arg	ysi
Leu	Val 210		Ile	Ile	Ser	Lys 215		Ser	Gln	His	Val 220	-	Tyr	Leu	Thr
Phe 225	Glu	Leu	Val	Leu	Met 230	Tyr	Сув	Asp	Val	Ile 235	Glu	Gly	Arg	Leu	Met 240
Thr	Glu	Thr	Ala	Met 245		Ile	Asp	Ala	Arg 250	Tyr	Thr	Glu	Leu	Leu 255	_
Arg	Val	Arg	Tyr 260	Met	Trp	Lys	Leu	11e 265	Asp	Gly	Phe	Phe	Pro 270		Leu
Gly	Asn	Pro 275	Thr	Tyr	Gln	Ile	Val 280	Ala	Met	Leu	Glu	Pro 285		Ser	Leu
Ala	Tyr 290	Leu	Gln	Leu	Arg	Asp 295	Ile	Thr	Val	Glu	Leu 300	_	Gly	Ala	Phe
Leu 305	Asn	His	Сув	Phe	Thr 310	Glu	Ile	His	Авр	Val 315	Leu	Авр	Gln	Asn	Gly 320
Phe	Ser	Авр	Glu	Gly 325	Thr	Tyr	His	Glu	Leu 330	Ile	Glu	Ala	Leu	<b>Авр</b> 335	Tyr
Ile	Phe	Ile	Thr 340	Asp	Asp	Ile	His	Leu 345	Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
Phe	Arg	Ser 355	Phe	Gly	His	Pro	Arg 360	Leu	Glu	Ala	Val	Thr 365	Ala	Ala	Glu
Asn	Val 370	Arg	Lys	Tyr	Met	Asn 375	Gln	Pro	Lys	Val	11e 380	Val	Tyr	Glu	Thr
Leu 385	Met	Lys	Gly	His	Ala 390	Ile	Phe	Сув	Gly	Ile 395	Ile	Ile	Asn	Gly	Tyr 400
Arg	qaƙ	Arg	His	Gly 405	Gly	Ser	Trp	Pro	Pro 410	Leu	Thr	Leu	Pro	Leu 415	His
Ala	Ala	Asp	Thr 420	Ile	Arg	Asn	Ala	Gln 425	Ala	Ser	Gly	Glu	Gly 430	Leu	Thr
His	Glu	Gln 435	Сув	Val	Asp	Asn	Trp 440	Lys	Ser	Phe	Ala	Gly 445	Val	Lys	Phe
Gly	Cys 450	Phe	Met	Pro	Leu	Ser 455	Leu	Asp	Ser	Asp	Leu 460	Thr	Met	Tyr	Leu

Lys 465	Asp	Lys	Ala	Leu	Ala 470	Ala	Leu	Gln	Arg	475		Asp	Ser	Val	Tyr 480
Pro	Lys	: Glu	Phe	Leu 485	Arg	Tyr	Asp	Pro	Pro 490		Gly	Thr	Gly	Ser 495	Arg
Arg	Leu	Val	<b>Asp</b> 500	Val	Phe	Leu	Asn	Asp 505		Ser	Phe	Asp	Pro 510		Asp
Val	Ile	Met 515	Tyr	Val	Val	Ser	Gly 520	Ala	Tyr	Leu	His	Asp 525		Glu	Phe
Asn	<b>Leu</b> 530	Ser	Tyr	Ser	Leu	Lys 535	Glu	Lys	Glu	Ile	Lys 540		Thr	Gly	Arg
Leu 545	Phe	Ala	Lys	Met	Thr 550	Tyr	Lys	Met	Arg	Ala 555	Сув	Gln	Val	Ile	Ala 560
Glu	Asn	Leu	Ile	Ser 565	Asn	Gly	Ile	Gly	Lys 570	Tyr	Phe	Lys	Авр	Asn 575	Gly
Met	Ala	Lys	Asp 580	Glu	His	Asp	Leu	Thr 585	Lys	Ala	Leu	His	Thr 590	Leu	Ala
Val	Ser	Gly 595	Val	Pro	Lys	Asp	Leu 600	Lys	Glu	Ser	His	Arg 605	Gly	Gly	Pro
Val	Leu 610	Lys	Thr	Tyr	Ser	Arg 615	Ser	Pro	Val	His	Thr 620	Ser	Thr	Arg	Asn
Val 625	Arg	Ala	Ala	Lys	Gly 630	Phe	Ile	Gly	Phe	Pro 635	Gln	Val	Ile	Arg	Gln 640
Asp	Gln	Авр	Thr	<b>А</b> вр 645	His	Pro	Glu	Asn	Met 650	Glu	Ala	Tyr	Glu	Thr 655	Val
Ser	Ala	Phe	11e 660	Thr	Thr	Авр	Leu	Lys 665	Lys	Tyr	Сув	Leu	Asn 670	Trp	Arg
Tyr	Glu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685	Ile	Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	Leu 700	Glu	Thr	Ser	Val
Leu 705	Tyr	Val	Ser	ysÞ	Pro 710	His	Сув	Pro	Pro	Asp 715	Leu	qsA	Ala	His	Ile 720
Pro	Leu	Tyr	Lys	Val 725	Pro	Asn	Asp	Gln	Ile 730	Phe	Ile	Lув	Tyr	Pro 735	Met

- 175 -

Gly Gly Ile Glu Gly Tyr Cys Gln Lys Leu Trp Thr Ile Ser Thr Ile Pro Tyr Leu Tyr Leu Ala Ala Tyr Glu Ser Gly Val Arg Ile Ala Ser Leu Val Gln Gly Asp Asn Gln Thr Ile Ala Val Thr Lys Arg Val Pro 775 Ser Thr Trp Pro Tyr Asn Leu Lys Lys Arg Glu Ala Ala Arg Val Thr Arg Asp Tyr Phe Val Ile Leu Arg Gln Arg Leu His Asp Ile Gly His His Leu Lys Ala Asn Glu Thr Ile Val Ser Ser His Phe Phe Val Tyr 825 Ser Lys Gly Ile Tyr Tyr Asp Gly Leu Leu Val Ser Gln Ser Leu Lys 835 Ser Ile Ala Arg Cys Val Phe Trp Ser Glu Thr Ile Val Asp Glu Thr 855 Arg Ala Ala Cys Ser Asn Ile Ala Thr Thr Met Ala Lys Ser Ile Glu 865 875 Arg Gly Tyr Asp Arg Tyr Leu Ala Tyr Ser Leu Asn Val Leu Lys Val 890 Ile Gln Gln Ile Leu Ile Ser Leu Gly Phe Thr Ile Asn Ser Thr Met 900 905 Thr Arg Asp Val Val Ile Pro Leu Leu Thr Asn Asn Asp Leu Leu Ile Arg Met Ala Leu Leu Pro Ala Pro Ile Gly Gly Met Asn Tyr Leu Asn 935 Met Ser Arg Leu Phe Val Arg Asn Ile Gly Asp Pro Val Thr Ser Ser 950 Ile Ala Asp Leu Lys Arg Met Ile Leu Ala Ser Leu Met Pro Glu Glu 970 Thr Leu His Gln Val Met Thr Gln Gln Pro Gly Asp Ser Ser Phe Leu 980 Asp Trp Ala Ser Asp Pro Tyr Ser Ala Asn Leu Val Cys Val Gln Ser Ile Thr Arg Leu Leu Lys Asn Ile Thr Ala Arg Phe Val Leu Ile His

- 176 -

	101	0				101	5				102	0			
Ser 1025	Pro 5	Asn	Pro	Met	Leu 103	Lys 0	Gly	Leu	Phe	His 103		Asp	Ser	Lys	Glu 1040
Glu	Asp	Glu	Gly	Leu 104	Ala 5	Ala	Phe	Leu	Met 105		Arg	His	Ile	Ile 105	Val 5
Pro	Arg	Ala	Ala 106	His O	Glu	Ile	Leu	Asp 106		Ser	Val	Thr	Gly 107		Arg
Glu	Ser	11e 107!	Ala 5	Gly	Met	Leu	Asp 108		Thr	Lys	Gly	Leu 108		Arg	Ala
Ser	Met 1090	Arg	Lys	Gly	Gly	Leu 109!		Ser	Arg	Val	Ile 110		Arg	Leu	Ser
Asn 1105	Tyr	Asp	туг	Glu	Gln 1110	Phe	Arg	Ala	Gly	Met 111!		Leu	Leu	Thr	Gly 1120
Arg	Lys	Arg	Asn	Val 1125	Leu	Ile	Asp	Lys	Glu 1130		Сув	Ser	Val	Gln 113	Leu 5
Ala	Arg	Ala	Leu 114(	Arg	Ser	His	Met	Trp 114	Ala 5	Arg	Leu	Ala	Arg 115		Arg
Pro	Ile	Tyr 1155	Gly	Leu	Glu	Val	Pro 1160		Val	Leu	Glu	Ser 116		Arg	Gly
His .	Leu 1170	Ile	Arg	Arg	His	Glu 1175		Сув	Val	Ile	Сув 118(		Сув	Gly	Ser
Val . 1185	Asn	Tyr	Gly	Trp	Phe 1190	Phe	Val	Pro	Ser	Gly 1195		Gln	Leu	Asp	Asp 1200
Cle /	Asp	Lys	Glu	Thr 1205	Ser	Ser	Leu	Arg	Val 1210		Tyr	Ile	Gly	Ser 1215	
Chr i	qaA		Arg 1220		Asp	Met		Leu 1225	Ala	Phe	Val	Arg	Ala 1230		Ser
lrg :	Ser	Leu 1235	Arg	Ser	Ala		Arg 1240		Ala	Thr	Val	Tyr 1245		Trp	Ala
Tyr (	Gly 1250	Asp .	Asp	Asp	Ser	Ser 1255	Trp	Asn	Glu		Trp 1260		Leu	Ala	Arg
31n 1 265	Arg	Ala .	Asn	Val	Ser 1270	Leu	Glu	Glu	Leu	Arg 1275		Ile	Thr	Pro	Ile 1280
er 1	Chr .	Ser '	Thr	Asn 1285	Leu .	Ala :	His .	Arg	Leu 1290		Asp	Arg		Thr 1295	

1295

- 177 -

- Val Lys Tyr Ser Gly Thr Ser Leu Val Arg Val Ala Arg Tyr Thr Thr 1300 1305 1310
- Ile Ser Asn Asp Asn Leu Ser Phe Val Ile Ser Asp Lys Lys Val Asp 1315 1320 1325
- Thr Asn Phe Ile Tyr Gln Gln Gly Met Leu Leu Gly Leu Gly Val Leu 1330 1335 1340
- Glu Thr Leu Phe Arg Leu Glu Lys Asp Thr Gly Ser Ser Asn Thr Val 1345 1350 1355 1360
- Leu His Leu His Val Glu Thr Asp Cys Cys Val Ile Pro Met Ile Asp 1365 1370 1375
- His Pro Arg Ile Pro Ser Ser Arg Lys Leu Glu Leu Arg Ala Glu Leu 1380 1385 1390
- Cys Thr Asn Pro Leu Ile Tyr Asp Asn Ala Pro Leu Ile Asp Arg Asp 1395 1400 1405
- Ala Thr Arg Leu Tyr Thr Gln Ser His Arg Arg His Leu Val Glu Phe 1410 1415 1420
- Val Thr Trp Ser Thr Pro Gln Leu Tyr His Ile Leu Ala Lys Ser Thr 1425 1430 1435 1440
- Ala Leu Ser Met Ile Asp Leu Val Thr Lys Phe Glu Lys Asp His Met 1445 1450 1455
- Asn Glu Ile Ser Ala Leu Ile Gly Asp Asp Asp Ile Asn Ser Phe Ile 1460 1465 1470
- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly 1475 1480 1485
- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565

- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Thr Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635. 1640 1645
- Met Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1655 1660
- Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys 1665 1670 1675 1680
- Gln Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala 1685 1690 1695
- Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn 1700 1705 1710
- Met Ser Ile Lys Ala Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu 1715 1720 1725
- Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly 1730 1735 1740
- Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn 1745 1750 1755 1760
- Ser Ser Ala Cys Tyr Lys Ala Val Glu Ile Ser Thr Leu Ile Arg Arg 1765 1770 1775
- Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly 1780 1785 1790
- Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe 1795 1800 1805
- Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu 1810 1815 1820
- Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val 1825 1830 1835 1840
- Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp

- 179 -

				184	5				185	0				185	5
Val	Gly	Ser	Val 186		Сув	Phe	Asn	Phe 186		Val	Ser	Asn	Ile 1870		Thr
Ser	Ser	Val 187	Gly 5	Phe	Ile	His	Ser 1880	-	Ile	Glu	Thr	Leu 1885		Asp	Lys
Авр	Thr 189		Glu	Lys	Leu	Glu 189		Leu	Ala	Ala	Ile 1900		Ser	Met	Ala
Leu 190		Leu	Gly	Lys	Ile 1910		Ser	Ile	Leu	Val 191		Lys	Leu	Met	Pro 1920
Phe	Ser	Gly	Asp	Phe 1925		Gln	Gly	Phe	Ile 1930		туг	Val	Gly	Ser 193	
Tyr	Arg	Glu	Val 1940		Leu	Val	Tyr	Pro 1945		Tyr	Ser	Asn	Phe 1950		Ser
Thr	Glu	Ser 195	Tyr 5	Leu	Val	Met	Thr 1960	-	Leu	Lys	Ala	Asn 1965		Leu	Met
Asn	Pro 1970		Lys	Ile	Lys	Gln 1975		Ile	Ile	Glu	Ser 1980		Val	Arg	Thr
Ser 1985		Gly	Leu	Ile	Gly 1990		Ile	Leu	Ser	Ile 1995	_	Gln	Leu	Ser	Сув 2000
1985	5		Leu		1990 Gly	)				1995 Arg	5				2000 Pro
1985	Gln	Ala		Val 2005 Leu	1990 Gly	Asp	Ala	Val	Ser 2010 Gln	1995 Arg	Gly	Asp	Ile	Asn 2015 Cys	2000 Pro
1985 Ile Thr	Gln Leu	Ala Lys	Ile Lys 2020 Asn	Val 2005 Leu	199( Gly ; Thr	Asp Pro	Ala Ile	Val Glu 2025 Cys	Ser 2010 Gln	1995 Arg Val	Gly	Asp Ile	Ile Asn 2030 His	Asn 2015 Cys	2000 Pro
1985 Ile Thr	Gln Leu Ala	Ala Lys Ile 2035 Ser	Ile Lys 2020 Asn	Val 2005 Leu )	Gly Gly Thr	Asp Pro Lys	Ala Ile Leu 2040	Val Glu 2025 Cys	Ser 2010 Gln Lys	Arg Val	Gly Leu Leu	Asp Ile Ile 2045 Leu	Ile Asn 2030 His	Asn 2015 Cys His	2000 Pro Gly Asp
1985 Ile Thr Leu Val	Gln Leu Ala Ala 2050	Lys Ile 2035	Lys 2020 Asn	Val 2005 Leu Gly Gln	Gly Thr Pro	Asp Pro Lys Gly 2055	Ala Ile Leu 2040 Leu	Val Glu 2025 Cys Leu	Ser 2010 Gln Lys Asn	Arg Val Glu	Gly Leu Leu Ile 2060	Ile Ile 2045 Leu	Ile Asn 2030 His	Asn 2015 Cys His Leu	2000 Pro Gly Asp
1985 Ile Thr Leu Val	Gln Leu Ala Ala 2050 Glu	Lys Ile 2035 Ser	Lys 2020 Asn Gly	Val 2005 Leu Gly Gln Arg	1990 Gly Thr Pro Asp Phe 2070	Asp Pro Lys Gly 2055	Ala Ile Leu 2040 Leu Asp	Val Glu 2025 Cys Leu Asn	Ser 2010 Gln Lys Asn	Arg Val Glu Ser Arg 2075	Gly Leu Leu Ile 2060	Ile Ile 2045 Leu	Ile Asn 2030 His Ile	Asn 2015 Cys His Leu	2000 Pro Gly Asp Tyr Met 2080
1985 Ile Thr Leu Val Arg 2065	Gln Leu Ala 2050 Glu His	Lys Ile 2035 Ser Leu Ala	Lys 2020 Asn Gly	Val 2005 Leu Gly Gln Arg Pro 2085	Gly  Thr  Pro  Asp  Phe 2070	Asp Pro Lys Gly 2055 Lys	Ala Ile Leu 2040 Leu Asp Val	Val Glu 2025 Cys Leu Asn	Ser 2010 Gln Lys Asn Gln Ser 2090	Arg Val Glu Ser Arg 2075	Gly Leu Leu 2060 Ser	Asp Ile Ile 2045 Leu Gln Arg	Asn 2030 His Ile Gln	Asn 2015 Cys His Leu Gly Leu 2095	2000 Pro Gly Asp Tyr Met 2080

- 180 -

Leu Ile Leu Asp Leu His Gln Asn Ile Phe Val Lys Asn Leu Ser Lys 2130 2135 2140

Ser Glu Lys Gln Ile Ile Met Thr Gly Gly Leu Lys Arg Glu Trp Val 2145 2150 2155 2160

Phe Lys Val Thr Val Lys Glu Thr Lys Glu Trp Tyr Lys Leu Val Gly 2165 2170 2175

Tyr Ser Ala Leu Ile Lys Asp 2180

# (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15894 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA (genomic)

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACCAAACAAA GTTGGGTAAG GATAGTTCAA TCAATGATCA TCTTCTAGTG CACTTAGGAT 60 TCAAGATCCT ATTATCAGGG ACAAGAGCAG GATTAGGGAT ATCCGAGATG GCCACACTTT 120 TAAGGAGCTT AGCATTGTTC AAAAGAAACA AGGACAAACC ACCCATTACA TCAGGATCCG 180 GTGGAGCCAT CAGAGGAATC AAACACATTA TTATAGTACC AATCCCTGGA GATTCCTCAA 240 TTACCACTCG ATCCAGACTT CTGGACCGGT TGGTGAGGTT AATTGGAAAC CCGGATGTGA 300 GCGGGCCCAA ACTAACAGGG GCACTAATAG GTATATTATC CTTATTTGTG GAGTCTCCAG 360 GTCAATTGAT TCAGAGGATC ACCGATGACC CTGACGTTAG CATAAGGCTG TTAGAGGTTG 420 TCCAGAGTGA CCAGTCACAA TCTGGCCTTA CCTTCGCATC AAGAGGTACC AACATGGAGG 480 ATGAGGCGGA CCAATACTTT TCACATGATG ATCCAATTAG TAGTGATCAA TCCAGGTTCG 540 GATGGTTCGG GAACAAGGAA ATCTCAGATA TTGAAGTGCA AGACCCTGAG GGATTCAACA 600 TGATTCTGGG TACCATCCTA GCCCAAATTT GGGTCTTGCT CGCAAAGGCG GTTACGGCCC 660 CAGACACGGC AGCTGATTCG GAGCTAAGAA GGTGGATAAA GTACACCCAA CAAAGAAGGG 720

- 181 -

TAGTTGGTGA	ATTTAGATTG	GAGAGAAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC	CTTACGCCGA	TTCATGGTCG	CTCTAATCCT	GGATATCAAG	AGAACACCCG	840
GAAACAAACC	CAGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATC	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCCTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGGT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAC	CTTTACCAGC	1020
aaatggggga	AACTGCACCC	TACATGGTAA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATC	ATACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG	AGGTTTGAAC	TTTGGCCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT	GGTAAGGAGG	TCAGCTGGAA	AGGTCAGTTC	CACATTGGCA	TCTGAACTCG	1260
GTATCACTGC	CGAGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCATACT	ACTGAGGACA	1320
AGATCAGTAG	AGCGGTTGGA	CCCAGACAAG	CCCAAGTATC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA	GCTACCGAGA	TTGGGGGGCA	AGGAAGATAG	GAGGGTCAAA	CAGAGTCGAG	1440
GAGAAGCCAG	GGAGAGCTAC	AGAGAAACCG	GGCCCAGCAG	AGCAAGTGAT	GCGAGAGCTG	1500
CCCATCTTCC	AACCGGCACA	CCCCTAGACA	TTGACACTGC	AACGGAGTCC	AGCCAAGATC	1560
CGCAGGACAG	TCGAAGGTCA	GCTGACGCCC	TGCTTAGGCT	GCAAGCCATG	GCAGGAATCT	1620
CGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTATAGTGTA	CAATGACAGA	AATCTTCTAG	1680
ACTAGGTGCG	AGAGGCCGAG	GGCCAGAACA	ACATCCGCCT	ACCATCCATC	ATTGTTATAA	1740
AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCCATCA	ACCATCCACT	CCCACGATTG	1800
GAGCCAATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	ATCGAGGAAG	CTATGGCAGC	ATGGTCAGAA	1920
ATATCAGACA	ACCCAGGACA	GGAGCGAGCC	ACCTGCAGGG	AAGAGAAGGC	AGGCAGTTCG	1980
GGTCTCAGCA	AACCATGCCT	CTCAGCAATT	GGATCAACTG	AAGGCGGTGC	ACCTCGCATC	2040
CGCGGTCAGG	GACCTGGAGA	GAGCGATGAC	GACGCTGAAA	CTTTGGGAAT	CCCCCAAGA	2100
					CAGCGGTGAA	
					TGATGGTGAT	
AGCACCCTCT	CAGGAGGAGA	CAATGAATCT	GAAAACAGCG	ATGTGGATAT	TGGCGAACCT	2280

- 182 -

GATACCGAGG GATATGCTAT CACTGACCGG GGATCTGCTC CCATCTCTAT GGGGTTCAGG	2340
GCTTCTGATG TTGAAACTGC AGAAGGAGGG GAGATCCACG AGCTCCTGAG ACTCCAATCC	2400
AGAGGCAACA ACTITCCGAA GCTTGGGAAA ACTCTCAATG TTCCTCCGCC CCCGGACCCC	2460
GGTAGGGCCA GCACTTCCGG GACACCCATT AAAAAGGGCA CAGACGCGAG ATTAGCCTCA	2520
TTTGGAACGG AGATCGCGTC TTTATTGACA GGTGGTGCAA CCCAATGTGC TCGAAAGTCA	2580
CCCTCGGAAC CATCAGGGCC AGGTGCACCT GCGGGGAATG TCCCCGAGTG TGTGAGCAAT	2640
GCCGCACTGA TACAGGAGTG GACACCCGAA TCTGGTACCA CAATCTCCCC GAGATCCCAG	2700
ANTANTGANG ANGGGGGNGN CTATTNTGNT GATGNGCTGT TCTCTGATGT CCANGATATT	2760
AAAACAGCCT TGGCCAAAAT ACACGAGGAT AATCAGAAGA TAATCTCCAA GCTAGAATCA	2820
CTGCTGTTAT TGAAGGGAGA AGTTGAGTCA ATTAAGAAGC AGATCAACAG GCAAAATATC	2880
AGCATATCCA CCCTGGAAGG ACACCTCTCA AGCATCATGA TCGCCATTCC TGGACTTGGG	2940
AAGGATCCCA ACGACCCCAC TGCAGATGTC GAAATCAATC CCGACTTGAA ACCCATCATA	3000
GGCAGAGATT CAGGCCGAGC ACTGGCCGAA GTTCTCAAGA AACCCGTTGC CAGCCGACAA	3060
CTCCAAGGAA TGACAAATGG ACGGACCAGT TCCAGAGGAC AGCTGCTGAA GGAATTTCAG	3120
CTAAAGCCGA TCGGGAAAAA GATGAGCTCA GCCGTCGGGT TTGTTCCTGA CACCGGCCCT	3180
GCATCACGCA GTGTAATCCG CTCCATTATA AAATCCAGCC GGCTAGAGGA GGATCGGAAG	3240
CGTTACCTGA TGACTCTCCT TGATGATATC AAAGGAGCCA ATGATCTTGC CAAGTTCCAC	3300
CAGATGCTGA TGAAGATAAT AATGAAGTAG CTACAGCTCA ACTTACCTGC CAACCCCATG	3360
CCAGTCGACC CAACTAGTAC AACCTAAATC CATTATAAAA AACTTAGGAG CAAAGTGATT	3420
GCCTCCCAAG GTCCACAATG ACAGAGACCT ACGACTTCGA CAAGTCGGCA TGGGACATCA	3480
AAGGGTCGAT CGCTCCGATA CAACCCACCA CCTACAGTGA TGGCAGGCTG GTGCCCCAGG	3540
TCAGAGTCAT AGATCCTGGT CTAGGCGACA GGAAGGATGA ATGCTTTATG TACATGTTTC	3600
TGCTGGGGGT TGTTGAGGAC AGCGATTCCC TAGGGCCTCC AATCGGGCGA GCATTTGGGT	3660
TCCTGCCCTT AGGTGTTGGC AGATCCACAG CAAAGCCCGA AAAACTCCTC AAAGAGGCCA	3720
CTGAGCTTGA CATAGTTGTT AGACGTACAG CAGGGCTCAA TGAAAAACTG GTGTTCTACA	3780
ACAACACCCC ACTAACTCTC CTCACACCTT GGAGAAAGGT CCTAACAACA GGGAGTGTCT	3840

- 183 -

TCAACGCAAA	CCAAGTGTGC	AATGCGGTTA	ATCTGATACC	GCTCGATACC	CCGCAGAGGT	3900
TCCGTGTTGT	TTATATGAGC	ATCACCCGTC	TTTCGGATAA	CGGGTATTAC	ACCGTTCCTA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	GGCGATAGGC	CCTGGGAAGA	TCATCGACAA	TACAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	GAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCTTATGT	TACCCGCTGA	TGGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380
AAGAATTCCG	CATTTACGAC	GACGTGATCA	TAAATGATGA	CCAAGGACTA	TTCAAAGTTC	4440
TGTAGACCGT	AGTGCCCAGC	AATGCCCGAA	AACGACCCCC	CTCACAATGA	CAGCCAGAAG	4500
GCCCGGACAA	AAAAGCCCCC	TCCGAAAGAC	TCCACGGACC	AAGCGAGAGG	CCAGCCAGCA	4560
GCCGACGGCA	AGCGCGAACA	CCAGGCGGCC	CCAGCACAGA	ACAGCCCTGA	CACAAGGCCA	4620
CCACCAGCCA	CCCCAATCTG	CATCCTCCTC	GTGGGACCCC	CGAGGACCAA	CCCCCAAGGC	4680
TGCCCCCGAT	CCAAACCACC	AACCGCATCC	CCACCACCCC	CGGGAAAGAA	ACCCCCAGCA	4740
ATTGGAAGGC	CCCTCCCCCT	CTTCCTCAAC	ACAAGAACTC	CACAACCGAA	CCGCACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGATCCTCTC	TCCCCGGCAA	4860
ACTARACARA	ACTTAGGGCC	AAGGAACATA	CACACCCAAC	AGAACCCAGA	CCCCGGTCCA	4920
				CCAACCAATC		4980
				CCCAGCACCC		5040
				GGGCCGACAG		5100
					AGACCACCCT	5160
					CAACCCGCGC	5220
					GAGCGATCCC	5280
					GTCCCCCGGT	5340
CTCCTCCTCT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CACCCGACGA	CACTCAACTC	5400

- 184 -

CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGAAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTAG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGGTT	5760
CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCGG	GAGTAGTCCT	GGCAGGTGCG	5820
GCCCTAGGCG	TTGCCACAGC	TGCTCAGATA	ACAGCCGGCA	TTGCACTTCA	CCAGTCCATG	5880
CTGAACTCTC	AAGCCATCGA	CAATCTGAGA	GCGAGCCTGG	AAACTACTAA	TCAGGCAATT	5940
GAGACAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
ATCAATAATG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTCGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCCAGTTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	TGGAGGAGAC	6180
ATCAATAAGG	TGTTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCGGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
AGTATAGCCT	ATCCGACGCT	GTCCGAGATT	AAGGGGGTGA	TTGTCCACCG	GCTAGAGGGG	6360
GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420
CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGGG	6540
TACACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGACGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATCATTGG	AGAGGTTGGA	CGTAGGGACA	6840
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	aattgttgga	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTATCG	AGCACTAGCA	TAGTCTACAT	CCTGATTGCA	6960

- 185 -

GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	· 7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACGGGA	7080
ACATCAAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCACCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATCAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	agtaggatag	TCATTAACAG	AGAACATCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTTG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
AATCTCTGAC	AAGATTAAAT	TCCTTAATCC	GGATAGGGAG	TACGACTTCA	GAGATCTCAC	7680
TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
GGCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
ATTCTCAAAC	ATGTCGCTGT	CCCTGTTAGA	CTTGTATTTA	GGTCGAGGTT	ACAATGTGTC	7920
ATCTATAGTC	ACTATGACAT	CCCAGGGAAT	GTATGGGGGA	ACTTACCTAG	TGGAAAAGCC	7980
TAATCTGAGC	AGCAAAAGGT	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040
AGGTGTTATC	AGAAATCCGG	GTTTGGGGGC	TCCGGTGTTC	CATATGACAA	ACTATCTTGA	8100
GCAACCAGTC	AGTAATGATC	TCAGCAACTG	TATGGTGGCT	TTGGGGGAGC	TCAAACTCGC	8160
AGCCCTTTGT	CACGGGGAAG	ATTCTATCAC	AATTCCCTAT	CAGGGATCAG	GGAAAGGTGT	8220
CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCTG	GAAATCCCCA	ACCGACATGC	AATCCTGGGT	8280
CCCCTTATCA	ACGGATGATC	CAGTGATAGA	CAGGCTTTAC	CTCTCATCTC	ACAGAGGTGT	8340
					ACAAGTTGCG	
					GCGAGAATCC	
CGAGTGGGCA	CCATTGAAGG	ATAACAGGAT	TCCTTCATAC	GGGGTCTTGT	CTGTTGATCT	8520

- 186 -

GAGTCTGACA GTTGAGCTTA AAATCAAAAT TGCTTCGGGA TTCGGGCCAT TGATCACACA	
CGGTTCAGGG ATGGACCTAT ACAAATCCAA CCACAACAAT GTGTATTGGC TGACTATCCC	8640
GCCAATGAAG AACCTAGCCT TAGGTGTAAT CAACACATTG GAGTGGATAC CGAGATTCAA	8700
GGTTAGTCCC TACCTCTTCA CTGTCCCAAT TAAGGAAGCA GGCGAAGACT GCCATGCCCC	8760
AACATACCTA CCTGCGGAGG TGGATGGTGA TGTCAAACTC AGTTCCAATC TGGTGATTCT	8820
ACCTGGTCAA GATCTCCAAT ATGTTTTGGC AACCTACGAT ACTTCCAGGG TTGAACATGC	8880
TGTGGTTTAT TACGTTTACA GCCCAAGCCG CTCATTTTCT TACTTTTATC CTTTTAGGTT	8940
GCCTATAAAG GGGGTCCCCA TCGAATTACA AGTGGAATGC TTCACATGGG ACCAAAAACT	9000
CTGGTGCCGT CACTTCTGTG TGCTTGCGGA CTCAGAATCT GGTGGACATA TCACTCACTC	9060
TGGGATGGTG GGCATGGGAG TCAGCTGCAC AGTCACCCGG GAAGATGGAA CCAATCGCAG	9120
ATAGGGCTGC TAGTGAACCA ATCACATGAT GTCACCCAGA CATCAGGCAT ACCCACTAGT	9180
GTGAAATAGA CATCAGAATT AAGAAAAACG TAGGGTCCAA GTGGTTCCCC GTTATGGACT	9240
CGCTATCTGT CAACCAGATC TTATACCCTG AAGTTCACCT AGATAGCCCG ATAGTTACCA	9300
ATAAGATAGT AGCCATCCTG GAGTATGCTC GAGTCCCTCA CGCTTACAGC CTGGAGGACC	9360
CTACACTGTG TCAGAACATC AAGCACCGCC TAAAAAACGG ATTTTCCAAC CAAATGATTA	9420
TAAACAATGT GGAAGTTGGG AATGTCATCA AGTCCAAGCT TAGGAGTTAT CCGGCCCACT	9480
CTCATATTCC ATATCCAAAT TGTAATCAGG ATTTATTTAA CATAGAAGAC AAAGAGTCAA	9540
CGAGGAAGAT CCGTGAACTC CTCAAAAAGG GGAATTCGCT GTACTCCAAA GTCAGTGATA	9600
AGGTTTTCCA ATGCTTAAGG GACACTAACT CACGGCTTGG CCTAGGCTCC GAATTGAGGG	9660
AGGACATCAA GGAGAAAGTT ATTAACTTGG GAGTTTACAT GCACAGCTCC CAGTGGTTTG	9720
AGCCCTTTCT GTTTTGGTTT ACAGTCAAGA CTGAGATGAG GTCAGTGATT AAATCACAAA	9780
CCCATACTTG CCATAGGAGG AGACACACA CTGTATTCTT CACTGGTAGT TCAGTTGAGT	9840
TGCTAATCTC TCGTGACCTT GTTGCTATAA TCAGTAAAGA GTCTCAACAT GTATATTACC	9900
TGACATTTGA ACTGGTTTTG ATGTATTGTG ATGTCATAGA GGGGAGGTTA ATGACAGAGA	9960
CCGCTATGAC TATTGATGCT AGGTATACAG AGCTTCTAGG AAGAGTCAGA TACATGTGGA	10020
AACTGATAGA TGGTTTCTTC CCTGCACTCG GGAATCCAAC TTATCAAATT GTAGCCATGC	10080

- 187 -

TGGAGCCTCT	TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTAA	CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
ATGAAGGTAC	TTATCATGAG	TTAACTGAAG	CTCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGAC	AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGGC	TGCTGAAAAT	GTTAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGAT	GAAAGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAGG	CAGTTGGCCA	CCGCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
					TGGAAATCTT	10560
					CTGACAATGT	10620
ACCTAAAGGA	CAAGGCACTT	GCTGCTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680
				GAGGCTTGTA		10740
				TGTTGTAAGT		10800
				AAAGGAGATC		10860
				CCAAGTGATT		10920
				GATGGCCAAG		10980
ATTTGACTAA						11040
				CCCAGTCCAC		11100
				AGTAATTCGG		11160
ACACTGATCA	TCCGGAGAAT	ATGGAAGCTT	ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAGAA						11280
TAAATGAGAT						11340
					ATCCCGTTAT	11400
ATAAAGTCCC						11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTATCTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGGAT						11580
TACCCAGCAC	ATGGCCCTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640

- 188 -

ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ATATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAA	ATTCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACAACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGGG	ATGAATTATC	12060
TGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCCTCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180
CACAACAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTGTATGTGT	CCAGAGCATC	ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTGA	12300
TCCATAGTCC	AAACCCAATG	TTAAAAGGAT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGGGACTGGC	GGCATTCCTC	ATGGACAGGC	ATATTATAGT	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATCA	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTG	GATACCACAA	12480
AAGGCTTGAT	TCGAGCCAGC	ATGAGGAAGG	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATTA	TGACTATGAA	CAATTCAGAG	CAGGGATGGT	GCTATTGACA	GGAAGAAAGA	12600
GAAATGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCGAGAGCT	CTAAGAAGCC	12660
ATATGTGGGC	GAGGCTAGCT	CGAGGACGGC	CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720
TAGAATCTAT	GCGAGGCCAC	CTTATTCGGC	GTCATGAGAC	ATGTGTCATC	TGCGAGTGTG	12780
GATCAGTCAA	CTACGGATGG	TTTTTTGTCC	CCTCGGGTTG	CCAACTGGAT	GATATTGACA	12840
AGGAAACATC	ATCCTTGAGA	GTCCCATATA	TTGGTTCTAC	CACTGATGAG	AGAACAGACA	12900
TGAAGCTTGC	CTTCGTAAGA	GCCCCAAGTC	GATCCTTGCG	ATCTGCTGTT	AGAATAGCAA	12960
CAGTGTACTC	ATGGGCTTAC	GGTGATGATG	ATAGCTCTTG	GAACGAAGCC	TGGTTGTTGG	13020
CTAGGCAAAG	GGCCAATGTG	AGCCTGGAGG	AGCTAAGGGT	GATCACTCCC	ATCTCAACTT	13080
CGACTAATTT	AGCGCATAGG	TTGAGGGATC	GTAGCACTCA	AGTGAAATAC	TCAGGTACAT	13140
CCCTTGTCCG	AGTGGCGAGG	TATACCACAA	TCTCCAACGA	CAATCTCTCA	TTTGTCATAT	13200

- 189 -

CAGATAAGAA	GGTTGATACT	AACTTTATAT	ACCAACAAGG	AATGCTTCTA	GGGTTGGGTG	13260
TTTTAGAAAC	ATTGTTTCGA	CTCGAGAAAG	ATACCGGATC	ATCTAACACG	GTATTACATC	13320
TTCACGTCGA	AACAGATTGT	TGCGTGATCC	CGATGATAGA	TCATCCCAGG	ATACCCAGCT	13380
CCCGCAAGCT	AGAGCTGAGG	GCAGAGCTAT	GTACCAACCC	ATTGATATAT	GATAATGCAC	13440
CTTTAATTGA	CAGAGATGCA	ACAAGGCTAT	ACACCCAGAG	CCATAGGAGG	CACCTTGTGG	13500
AATTTGTTAC	ATGGTCCACA	CCCCAACTAT	ATCACATTTT	AGCTAAGTCC	ACAGCACTAT	13560
CTATGATTGA	CCTGGTAACA	AAATTTGAGA	AGGACCATAT	GAATGAAATT	TCAGCTCTCA	13620
TAGGGGATGA	CGATATCAAT	AGTTTCATAA	CTGAGTTTCT	GCTCATAGAG	CCAAGATTAT	13680
TCACTATCTA	CTTGGGCCAG	TGTGCGGCCA	TCAATTGGGC	ATTTGATGTA	CATTATCATA	13740
GACCATCAGG	GAAATATCAG	ATGGGTGAGC	TGTTGTCATC	GTTCCTTTCT	AGAATGAGCA	13800
AAGGAGTGTT	TAAGGTGCTT	GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
			GTCCTTCACT			13920
			ATATGACCTA			13980
			GTGAAAGCGA			14040
			GTGTTCTGGC			14100
			CGGTAGAGAA			14160
			CAGGATCTTC			14220
			ATCTCCGGCG			14280
			ACGCCCTCGC			14340
			TGAGCATCAA			14400
			ACACAAGCAA			14460
					AACTCATCTG	14520
			TAATTAGGAG			14580
					GAGATACTTA	14640
			TTTCCGCCAA TTGTCGAACA			14700
	CIMICCICC	GANGI I GGCC	IIGICGAACA	CAGAATGGGA	GTAGGTAATA	14760

- 190 -

T	TGTCAAAGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	GTAGATTGCT	14820
T	CAATTTCAT	AGTTAGTAAT	ATCCCTACCT	CTAGTGTGGG	GTTTATCCAT	TCAGATATAG	14880
A	GACCTTGCC	TGACAAAGAT	ACTATAGAGA	AGCTAGAGGA	ATTGGCAGCC	ATCTTATCGA	14940
T	GGCTCTGCT	CCTGGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
G	GGATTTTGT	TCAGGGATTT	ATAAGTTATG	TAGGGTCTCA	TTATAGAGAA	GTGAACCTTG	15060
T	ATACCCTAG	ATACAGCAAC	TTCATCTCTA	CTGAATCTTA	TTTGGTTATG	ACAGATCTCA	15120
A	GGCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	GATAATTGAA	TCATCTGTGA	15180
G	GACTTCACC	TGGACTTATA	GGTCACATCC	TATCCATTAA	GCAACTAAGC	TGCATACAAG	15240
C	aattgtggg	AGACGCAGTT	AGTAGAGGTG	ATATCAATCC	TACTCTGAAA	AAACTTACAC	15300
C.	PATAGAGCA	GGTGCTGATC	AATTGCGGGT	TGGCAATTAA	CGGACCTAAG	CTGTGCAAAG	15360
A	ATTGATCCA	CCATGATGTT	GCCTCAGGGC	AAGATGGATT	GCTTAATTCT	ATACTCATCC	15420
T	CTACAGGGA	GTTGGCAAGA	TTCAAAGACA	ACCAAAGAAG	TCAACAAGGG	ATGTTCCACG	15480
C.	TACCCCGT	ATTGGTAAGT	AGCAGGCAAC	GAGAACTTAT	ATCTAGGATC	ACCCGCAAAT	15540
T	CTGGGGGCA	CATTCTTCTT	TACTCCGGGA	ACAAAAAGTT	GATAAATAAG	TTTATCCAGA	15600
A?	<b>ICTCAAGTC</b>	CGGCTATCTG	ATACTAGACT	TACACCAGAA	TATCTTCGTT	AAGAATCTAT	15660
C	CAAGTCAGA	GAAACAGATT	ATTATGACGG	GGGGTTTGAA	ACGTGAGTGG	GTTTTTAAGG	15720
T	AACAGTCAA	GGAGACCAAA	GAATGGTATA	AGTTAGTCGG	ATACAGTGCC	CTGATTAAGG	15780
A	CTAATTGGT	TGAACTCCGG	AACCCTAATC	CTGCCCTAGG	TGGTTAGGCA	TTATTTGCAA	15840
T	ATATTAAAG	AAAACTTTGA	AAATACGAAG	TTTCTATTCC	CAGCTTTGTC	TGGT	15894

- (2) INFORMATION FOR SEQ ID NO:12:
  - (i) SEQUENCE CHARACTERISTICS:
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

(A) LENGTH: 2183 amino acids

- 191 -

	Met 1	Asp	Ser	Leu	Ser 5	Val	Asn	Gln	Ile	Leu 10	Tyr	Pro	Glu	Val	His 15	Le
	<b>Asp</b>	Ser	Pro	Ile 20	Val	Thr	Asn	Lys	Ile 25	Val	Ala	Ile	Leu	Glu 30	Tyr	Ala
	Arg	Val	Pro 35	His	Ala	Tyr	Ser	Leu 40	Glu	Asp	Pro	Thr	Leu 45	Сув	Gln	Ası
	Ile	Lys 50	His	Arg	Leu	Lys	Asn 55	Gly	Phe	Ser	Asn	Gln 60	Met	Ile	Ile	Хei
	65					70				Ser	75					80
					85					Сув 90					95	
				100					105	Ile				110		
			115					120		Авр			125			•
		130					135			Gly		140				_
	145					150				Val	155					160
					165					Thr 170					175	
				180					185	Сув				190		
			195					200		Glu			205			
		210					215			Gln		220		-		
	225					230				Val	235		-			240
					245					Arg 250					255	
•	Arg	Val	Arg	Tyr 260	Met	Trp	Lys		Ile 265	Asp	Gly	Phe	Phe	Pro 270	Ala	Leu

- 192 -

GTÅ	ABI	275	Thr	Тух	c Gln	ı Ile	280		a Met	Lev	ı Glı	285		ı Ser	. Le
Ala	1yr 290	Leu	Glr	Leu	Arg	295	ıle i	Thr	· Val	Glu	300		g Gly	Ala	Ph
Leu 305	Asn	His	Сув	Phe	310	Glu	ılle	His	a Ası	Val 315		а Авр	Gln	Asn	32
Phe	Ser	yab	Glu	Gly 325	Thr	Tyr	His	Glu	330	Thr	Glu	Ala	Leu	Asp 335	_
Ile	Phe	Ile	340	ysb	Asp	Ile	His	345		Gly	Glu	Ile	Phe 350		Ph
		355					360			Ala		365			
	370					375				Val	380				
385					390					Ile 395					400
				405					410					415	
			420					425		Ser			430		
		435					440			Phe		445			
	450					455				Asp	460				
465					470					Glu 475					480
				485					490	Lys				495	
			500					505		Ser			510		
		515					520			Leu		525			
	530					535				Ile	540				
Leu	Phe	Ala	Lys	Met	Thr	Tyr	Lys	Met	Arg	Ala	Сув	Gln	Val	Ile	Ala

- 193 -

545					550					555	<b>i</b>				560
Glů	Asn	Leu	Ile	Ser 565	Asn	Gly	Ile	Gly	Lys 570		Phe	Lys	Asp	575	_
Met	Ala	Lys	<b>Авр</b> 580		His	Asp	Leu	Thr 585		Ala	Leu	His	590		Ala
Val	Ser	Gly 595		Pro	Lys	Asp	Leu 600		Glu	Ser	His	Arg 605	_	Gly	Pro
Val	Leu 610	Lys	Thr	Tyr	Ser	Arg 615		Pro	Val	His	Thr 620		Thr	Arg	Asn
Val 625	Arg	Ala	Ala	Lys	Gly 630	Phe	Ile	Gly	Phe	Pro 635		Val	Ile	Arg	Gln 640
Asp	Gln	Asp	Thr	Asp 645	His	Pro	Glu	Asn	Met 650	Glu	Ala	Tyr	Glu	Thr 655	
Ser	Ala	Phe	11e 660	Thr	Thr	Asp	Leu	<b>Lys</b> 665	Lys	Tyr	Сув	Leu	Asn 670	Trp	λrg
Tyr	Gļu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685	Ile	Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	Leu 700	Glu	Thr	Ser	Val
<b>Leu</b> 705	Tyr	Val	Ser	Asp	Pro 710	His	Сув	Pro	Pro	Asp 715	Leu	Aap	Ala	His	Ile 720
				725	Pro				730					735	
			740		Tyr			745					750		
Pro	Tyr	Leu 755	Tyr	Leu	Ala	Ala	Tyr 760	Glu	Ser	Gly	Val	Arg 765	Ile	Ala	Ser
Leu	Val 770	Gln	Gly	Asp	Asn	Gln 775	Thr	Ile	Ala	Val	Thr 780	Lys	Arg	Val	Pro
9er 785	Thr	Trp	Pro	Tyr	Asn 790	Leu	Lys	Lys	Arg	Glu 795	Ala	Ala	Arg	Val	Thr 800
				805	Ile				810					815	
His	Leu	Lys	Ala 820	Asn	Glu	Thr	Ile	Val 825	Ser	Ser	His	Phe	Phe 830	Val	Tyr

Ser	Lys	Gly 835	Ile	Tyr	Tyr	Asp	Gly 840	Leu	Leu	Val	Ser	Gln 845	Ser	Leu	Lys
Ser	11e 850	Ala	Arg	Сув	Val	Phe 855	Trp	Ser	Glu	Thr	Ile 860	Val	Asp	Glu	Thr
Arg 865	Ala	Ala	Сув	Ser	Asn 870	Ile	Ala	Thr	Thr	Met 875	Ala	Lys	Ser	Ile	Glu 880
Arg	Gly	Tyr	Asp	Arg 885	Тух	Leu	Ala	Tyr	Ser 890	Leu	Asn	Val	Leu	Lув 895	Val
Ile	Gln	Gln	Ile 900	Leu	Ile	Ser	Leu	Gly 905	Phe	Thr	Ile	Asn	Ser 910	Thr	Met
	Arg	915					920					925			
	Met 930					935					940				
945	Ser				950					955					960
	Ala			965					970					975	
	Leu		980					985					990		
	Trp	995					1000	)				1005	5		
	Thr 1010	)				1015	5				1020	)			
1025					1030	)				1035	5				1040
Glu	Asp	Glu	Gly	Leu 1045		Ala	Phe	Leu	Met 105(	_	Arg	His	Ile	11e 1055	
	Arg		1060	)				1065	5				1070	)	
	Ser	1075	5				1080	)				1085	5		
Ser	Met 1090		Lys	Gly		Leu 1095		Ser	Arg		Ile 1100		Arg	Leu	Ser

- 195 -

Asn 110		Asp	Tyr	Glu	Gln 111		Arg	Ala	Gly	Met 111		Leu	Leu	Thr	Gly 1120
Arg	Lys	Arg	Asn	Val 112		Ile	Asp	Lys	Glu 113		Сув	Ser	Val	Gln 113	Leu
Ala	Arg	Ala	Leu 1140	-	Ser	His	Met	Trp 114		Arg	Leu	Ala	Arg 115		Arg
Pro	Ile	Tyr 115		Leu	Glu	Val	Pro 1160	_	Val	Leu	Glu	Ser 116		Arg	Gly
His	Leu 1170		Arg	Arg	His	Glu 1179		Сув	Val	Ile	Cys 118		Сув	Gly	Ser
Val 118		Tyr	Gly	Trp	Phe 1190		Val	Pro	Ser	Gly 119		Gln	Leu	Asp	Авр 1200
Ile	Asp	Lys	Glu	Thr 1205		Ser	Leu	Arg	Val 121		Tyr	Ile	Gly	Ser 121	
Thr	Asp	Glu	Arg 1220		Asp	Met	Lys	Leu 122		Phe	Val	Arg	Ala 1230		Ser
Arg	Ser	Leu 1235	_	Ser	Ala	Val	Arg 1240		Ala	Thr	Val	Tyr 1245		Trp	Ala
Tyr	Gly 1250	_	Asp	Asp	Ser	Ser 1255	_	Asn	Glu	Ala	Trp 1260		Leu	Ala	Arg
Gln 126	_	Ala	Asn	Val	Ser 1270		Glu	Glu	Leu	Arg 1279		Ile	Thr	Pro	Ile 1280
Ser	Thr	Ser	Thr	Asn 1285		Ala	His	Arg	Leu 1290	_	Asp	Arg	Ser	Thr 1295	
Val	Lys	Tyr	Ser 1300	_	Thr	Ser	Leu	Val 130	_	Val	Ala	Arg	Tyr 1310		Thr
Ile	Ser	Asn 1315	_	Asn	Leu	Ser	Phe 1320		Ile	Ser	Asp	Lys 1325	_	Val	Asp
	Asn 1330		Il•	Tyr	Gln	Gln 1335	_	Met	Leu	Leu	Gly 1340		Gly	Val	Leu
Glu 134!		Leu	Phe	Arg	Leu 1350		Lys	Asp	Thr	Gly 1355		Ser	Asn	Thr	Val 1360
Leu	His	Leu	His	Val 1365		Thr	Asp	Сув	Сув 1370		Ile	Pro	Met	Ile 1375	_
Hic	Drc	A	TIO	Drc	Car	Car	A ===	Tare	Lan	Gl v	Low	3 ~~	31 e	C1	Lou

- 196 -

1380 1385 1390

- Cys Thr Asn Pro Leu Ile Tyr Asp Asn Ala Pro Leu Ile Asp Arg Asp 1395 1400 1405
- Ala Thr Arg Leu Tyr Thr Gln Ser His Arg Arg His Leu Val Glu Phe 1410 1415 1420
- Val Thr Trp Ser Thr Pro Gln Leu Tyr His Ile Leu Ala Lys Ser Thr 1425 1430 1435 1440
- Ala Leu Ser Met Ile Asp Leu Val Thr Lys Phe Glu Lys Asp His Met 1445 1450 1455
- Asn Glu Ile Ser Ala Leu Ile Gly Asp Asp Asp Ile Asn Ser Phe Ile 1460 1465 1470
- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly 1475 1480 1485
- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565
- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Thr Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635 1640 1645
- Met Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1660

- 197 -

Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys Gln Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn Met Ser Ile Lys Ala Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn Ser Ser Ala Cys Tyr Lys Ala Val Glu Ile Ser Thr Leu Ile Arg Arg Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp Val Gly Ser Val Asp Cys Phe Asn Phe Ile Val Ser Asn Ile Pro Thr Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asp Lys Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala Leu Leu Cly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro

Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser His

#### - 198 -

- Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser 1940 1945 1950
- Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965
- Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1980
- Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000
- Ile Gln Ala Ile Val Gly Asp Ala Val Ser Arg Gly Asp Ile Asn Pro 2005 2010 2015
- Thr Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Asn Cys Gly 2020 2025 2030
- Leu Ala Ile Asn Gly Pro Lys Leu Cys Lys Glu Leu Ile His His Asp 2035 2040 2045
- Val Ala Ser Gly Gln Asp Gly Leu Leu Asn Ser Ile Leu Ile Leu Tyr 2050 2055 2060
- Arg Glu Leu Ala Arg Phe Lys Asp Asn Gln Arg Ser Gln Gln Gly Met 2065 2070 2075 2080
- Phe His Ala Tyr Pro Val Leu Val Ser Ser Arg Gln Arg Glu Leu Ile 2085 2090 2095
- Ser Arg Ile Thr Arg Lys Phe Trp Gly His Ile Leu Leu Tyr Ser Gly 2100 2105 2110
- Asn Lys Leu Ile Asn Lys Phe Ile Gln Asn Leu Lys Ser Gly Tyr 2115 2120 2125
- Leu Ile Leu Asp Leu His Gln Asn Ile Phe Val Lys Asn Leu Ser Lys 2130 2135 2140
- Ser Glu Lys Gln Ile Ile Met Thr Gly Gly Leu Lys Arg Glu Trp Val 2145 2150 2155 2160
- Phe Lys Val Thr Val Lys Glu Thr Lys Glu Trp Tyr Lys Leu Val Gly 2165 2170 2175
- Tyr Ser Ala Leu Ile Lys Asp 2180
- (2) INFORMATION FOR SEQ ID NO:13:
  - (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 15894 base pairs

- 199 -

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: RNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ACCAAACAAA GTTGGGTAAG GATAGTTCAA TCAATGATCA TTTTCTAGTG CACTTAGGAT 60 120 TCAAGATCCT ATTATCAGGG ACAAGAGCAG GATTAAGGAT ATCCGAGATG GCCACACTTT 180 TAAGGAGCTT AGCATTGTTC AAAAGAAACA AGGACAAACC ACCCATTACA TCAGGATCCG GTGGAGCCAT CAGAGGAATC AAACACATTA TTATAGTACC AATCCCTGGA GATTCCTCAA 300 TTACCACTCG ATCCAGACTT CTGGACCGGT TGGTCAGGTT AATTGGAAAC CCGGATGTGA GCGGGCCCAA ACTAACAGGG GCACTAATAG GTATATTATC CTTATTTGTG GAGTCTCCAG 360 GTCAATTGAT TCAGAGGATC ACCGATGACC CTGACGTTAG CATAAGGCTG TTAGAGGTTG 420 TCCAGAGTGA CCAGTCACAA TCTGGCCTTA CCTTCGCATC AAGAGGTACC AACATGGAGG 480 ATGAGGCGGA CCAATACTTT TCACATGATG ATCCAATTAG TAGTGATCAA TCCAGGTTCG 540 GATGGTTCGA GAACAAGGAA ATCTCAGATA TTGAAGTGCA AGACCCTGAG GGATTCAACA 600 660 TGATTCTGGG TACCATCCTA GCTCAAATTT GGGTCTTGCT CGCAAAGGCG GTTACGGCCC CAGACACGGC AGCTGATTCG GAGCTAAGAA GGTGGATAAA GTACACCCAA CAAAGAAGGG 720 780 TAGTTGGTGA ATTTAGATTG GAGAGAAAAT GGTTGGATGT GGTGAGGAAC AGGATTGCCG 840 AGGACCTCTC CTTACGCCGA TTCATGGTCG CTCTAATCCT GGATATCAAG AGAACACCCG 900 GAAACAAACC CAGGATTGCT GAAATGATAT GTGACATTGA TACATATATC GTAGAGGCAG GATTAGCCAG TTTTATCCTG ACTATTAAGT TTGGGATAGA AACTATGTAT CCTGCTCTTG 960 GACTGCATGA ATTTGCTGGT GAGTTATCCA CACTTGAGTC CTTGATGAAC CTTTACCAGC 1020 1080 AAATGGGGGA AACTGCACCC TACATGGTAA TCCTGGAGAA CTCAATTCAG AACAAGTTCA GTGCAGGATC ATACCCTCTG CTCTGGAGCT ATGCCATGGG AGTAGGAGTG GAACTTGAAA 1140 1200 ACTCCATGGG AGGTTTGAAC TTTGGCCGAT CTTACTTTGA TCCAGCATAT TTTAGATTAG GGCAAGAGAT GGTAAGGAGG TCAGCTGGAA AGGTCAGTTC CACATTGGCA TCTGAACTCG 1260

- 200 -

GTATCACTO	C CGAGGATGO	CA AGGCTTGTT	TT CAGAGATTO	C AATGCATA	CT ACTGAGGACA	1320
AGATCAGTA	G AGCGGTTGG	ga cccagacai	G CCCAAGTAT	C ATTTCTAC	C GGTGATCAAA	1380
GTGAGAATG	A GCTACCGAG	A TTGGGGGG	A AGGAAGATA	G GAGGGTCA	LA CAGAGTCGAG	1440
GAGAAGCCA	G GGAGAGCTA	C AGAGAAACC	G GGCCCAGCA	.G AGCAAGTGA	T GCGAGAGCTG	1500
					C AGCCAAGATC	1560
					G GCAGGAATCT	1620
					A AATCTTCTAG	1680
					C ATTGTTATAA	1740
					T CCCACGATTG	1800
					G CATCCGGGCT	1860
					C ATGGTCAGAA	1920
					C AGGCAGTTCG	1980
					ACCTCGCATC	2040
					CCCCCAAGA	2100
					CAGCGGTGAA	2160
					TGATGGTGAT	2220
					TGGCGAACCT	2280
					GGGGTTCAGG	2340
					ACTCCAATCC	2400
	ACTTTCCGAA					2460
					ATTAGCCTCA	2520
					TCGAAAGTCA	2580
					TGTGAGCAAT	2640
					GAGATCCCAG	2700
					CCAAGATATT	2760
	TGGCCAAAAT					2820

- 201 -

CTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAGAAGC	AGATCAACAG	GCAAAATATC	2880
AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TCGCCATTCC	TGGACTTGGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTC	GAAATCAATC	CCGACTTGAA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AACCCGTTGC	CAGCCGACAA	3060
CTCCAAGGAA	TGACAAATGG	ACGGACCAGT	TCCAGAGGAC	AGCTGCTGAA	GGAATTTCAG	3120
CTAAAGCCGA	TCGGGAAAAA	GATGAGCTCA	GCCGTCGGGT	TTGTTCCTGA	CACCGGCCCT	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGA	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ATGATCTTGC	CAAGTTCCAC	3300
CAGATGCTGA	TGAAGATAAT	AATGAAGTAG	CTACAGCTCA	ACTTACCTGC	CAACCCCATG	3360
CCAGTCGACC	CAACTAGTAC	AACCTAAATC	CATTATAAAA	AACTTAGGAG	CAAAGTGATT	3420
GCCTCCCAAG	TTCCACAATG	ACAGAGATCT	ACGACTTCGA	CAAGTCGGCA	TGGGACATCA	3480
AAGGGTTGAT	CGCTCCGATA	CAACCCACCA	CCTACAGTGA	TGGCAGGCTG	GTGCCCCAGG	3540
TCAGAGTCAT	AGATCCTGGT	CTAGGCGACA	GGAAGGATGA	ATGCTTTATG	TACATGTTTC	3600
TGCTGGGGGT	TGTTGAGGAC	AGCGATCCCC	TAGGGCCTCC	AATCGGGCGA	GCATTTGGGT	3660
CCCTGCCCTT	AGGTGTTGGC	AAATCCACAG	CAAAGCCCGA	AAAACTCCTC	AAAGAGGCCA	3720
CTGAGCTTGA	CATAGTTGTT	AGACGTACAG	CAGGGCTCAA	TGAAAAACTG	GTGTTCTACA	3780
ACAACACCCC	ACTAACTCTC	CTCACACCTT	GGAGAAAGGT	CCTAACAACA	GGGAGTGTCT	3840
TCAACGCAAA	CCAAGTGTGC	AGTGCGGTTA	ATCTGATACC	GCTCGATACC	CCGCAGAGGT	3900
TCCGTGTTGT	TTATATGAGC	ATCACCCGTC	TTTCGGATAA	CGGGTATTAC	ACCGTTCCTA	3960
GAAGAATGCT	GGAATTCAGA	TCGGTCAATG	CAGTGGCCTT	CAACCTGCTG	GTGACCCTTA	4020
GGATTGACAA	GGCGATAGGC	CCTGGGAAGA	TCATCGACAA	TACAGAGCAA	CTTCCTGAGG	4080
CAACATTTAT	GGTCCACATC	GGGAACTTCA	GGAGAAAGAA	GAGTGAAGTC	TACTCTGCCG	4140
ATTATTGCAA	AATGAAAATC	GAAAAGATGG	GCCTGGTTTT	TGCACTTGGT	GGGATAGGGG	4200
GCACCAGTCT	TCACATTAGA	AGCACAGGCA	AAATGAGCAA	GACTCTCCAT	GCACAACTCG	4260
GGTTCAAGAA	GACCTTATGT	TACCCGCTGA	TAGATATCAA	TGAAGACCTT	AATCGATTAC	4320
TCTGGAGGAG	CAGATGCAAG	ATAGTAAGAA	TCCAGGCAGT	TTTGCAGCCA	TCAGTTCCTC	4380

- 202 -

AAGAATTCC	G CATTTACGA	C GACGTGATC	A TAAATGATG	A CCAAGGACT	A TTCAAAGTTC	4440
TGTAGACCG!	r agtgcccag	C AATGCCCGAL	AACGACCCC	C CTCACAATG	A CAGCCAGAAG	4500
GCCCGGACAI	A AAAAGCCCC	C TCCGAAAGA	CTCCACGGAC	C AAGCGAGAG	G CCAGCCAGCA	4560
GCCGACGGC	AGCGCGAAC	A CCAGGCGGC	CCAGCACAG	A ACAGCCCTG	A TACAAGGCCA	4620
CCACCAGCC	CCCCAATCT	G CATCCTCCTC	GTGGGACCC	CGAGGACCA!	CCCCCAAGGC	4680
TGCCCCCGA1	CCAAACCAC	C AACCGCATCO	CCACCACCC	CGGGAAAGA	ACCCCCAGCA	4740
ATTGGAAGG	CCCTCCCCC	CTTCCTCAAC	CAAGAACTC	CACAACCGAR	CCGCACAAGC	4800
GACCGAGGTG	ACCCAACCGC	AGGCATCCGA	CTCCCTAGAC	AGATCCTCTC	TCCCCGGCAA	4860
					CCCCGGCCCA	4920
					CCGCCGGCTC	4980
					AACCATCGAC	5040
		CCCCAAAAA				5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CGGCAACCAA	ACCAGAACCC	AGACCACCCT	5160
GGGCCACCAG	CTCCCAGACT	CGGCCATCAC	CCCGCAGAAA	GGAAAGGCCA	CAACCCGCGC	5220
ACCCCAGCCC	CGATCCGGCG	GGGAGCCACC	CAACCCGAAC	CAGCACCCAA	GAGCGATCCC	5280
CGAAGGACCC	CCGAACCGCA	AAGGACATCA	GTATCCCACA	GCCTCTCCAA	GTCCCCCGGT	5340
CTCCTCCCCT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CACCCGACGA	CACTCAACTC	5400
CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	aagatagggg	TGGTAGGAAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTAG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
		AGATGCACTT				5760
		GAGACACAAG				5820
		TGCTCAGATA				5880
		Caatctgaga				5940

- 203 -

GAGGCAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
ATCAATAATG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTCGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCCAGCTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	TGGAGGAGAC	6180
ATCAATAAGG	TGTTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
AGTATAGCCT	ATCCGACGCT	GTCCGAGATT	AAGGGGTGA	TTGTCCACCG	GCTAGAGGGG	6360
GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420
CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGGG	6540
TCCACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TCACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGATGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATCATTGG	AGAGGTTGGA	CGTAGGGACA	6840
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	AATTGTTGGA	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTATCG	AGCACTAGCA	TAGTCTACAT	CCTGATTGCA	6960
GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACGGGA	7080
ACATCAAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCACCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	AGTAGGATAG	TCATTAACAG	AGAACATCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTTG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500

- 204 -

TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
CATCTCTGAC	AAGATTAAAT	TCCTTAATCC	GGATAGGGAG	TACGACTICA	GAGATCTCAC	7680
TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
GGCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
ATTCTCAAAC	ATGTCGCTGT	CCCTGTTAGA	CTTGTATTTA	GGTCGAGGTT	ACAATGTGTC	7920
ATCTATAGTC	ACTATGACAT	CCCAGGGAAT	GTATGGGGGA	ACTTACCTAG	TGGAAAAGCC	7980
TAATCTGAGC	AGCAAAAGGT	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040
AGGTGTTATC	AGAAATCCGG	GTTTGGGGGC	TCCGGTGTTC	CATATGACAA	ACTATCTTGA	8100
GCAACCAGCC	AGTAATGATC	TCAGCAACTG	TATGGTGGCT	TTGGGGGAGC	TCAAACTCGC	8160
AGCCCTTTGT	CACGGGGAAG	ATTCTATCAC	AATTCCCTAT	CAGGGATCAG	GGAAAGGTGT	8220
CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCTG	GAAATCCCCA	ACCGACATGC	AATCCTGGGT	8280
CCCCTTATCA	ACGGATGATC	CAGTGATAGA	CAGGCTTTAC	CTCTCATCTC	ACAGAGGTGT	8340
TATCGCTGAC	AATCAAGCAA	AATGGGCTGT	CCCGACAACA	CGAACAGATG	ACAAGTTGCG	8400
AATGGAGACA	TGCTTCCAAC	AGGCGTGTAA	GGGTAAAATC	CAAGCACTCT	GCGAGAATCC	8460
CGAGTGGGCA	CCATTGAAGG	ATAACAGGAT	TCCTTCATAC	GGGGTCTTGT	CTGTTGATCT	8520
GAGTCTGACA	GTTGAGCTTA	AAATCAAAAT	TGCTTCGGGA	TTCGGGCCAT	TGATCACACA	8580
CGGTTCAGGG	ATGGACCTAT	ACAAATCCAA	CCACAACAAT	GTGTATTGGC	TGACTATCCC	8640
GCCAATGAAG	AACCTAGCCT	TAGGTGTAAT	CAACACATTG	GAGTGGATAC	CGAGATTCAA	8700
GGTTAGTCCC	TACCTCTTCA	ATGTCCCAAT	TAAGGAAGCA	GGCGAAGACT	GCCATGCCCC	8760
AACATACCTA	CCTGCGGAGG	TGGATGGTGA	TGTCAAACTC	AGTTCCAATC	TGGTGATTCT	8820
ACCTGGTCAA	GATCTCCAAT	ATGTTTTGGC	AACCTACGAT	ACTTCCAGGG	TTGAACATGC	8880
TGTGGTTTAT	TACGTTTACA	GCCCAGGCCG	CTCATTTTCT	TACTTTTATC	CTTTTAGGTT	8940
GCCTATAAAG	GGGGTCCCCA	TCGAATTACA	AGTGGAATGC	TTCACATGGG	ACCAAAAACT	9000
CTGGTGCCGT	CACTTCTGTG	TGCTTGCGGA	CTCAGAATCT	GGTGGACATA	TCACTCACTC	9060

- 205 -

				· ·		
TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACCCGG	GAAGATGGAA	CCAATCGCAG	9120
ATAGGGCTGC	TAGTGAACCA	ATCTCATGAT	GTCACCCAGA	CATCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATT	AAGAAAAACG	TAGGGTCCAA	GTGGTTCCCC	GTTATGGACT	9240
CGCTATCTGT	CAACCAGATC	TTATACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
ATAAGATAGT	AGCCATCCTG	GAGTATGCTC	GAGTCCCTCA	CGCTTACAGC	CTGGAGGACC	9360
CTACACTGTG	TCAGAACATC	AAGCACCGCC	TAAAAAACGG	ATTTTCCAAC	CAAATGATTA	9420
TAAACAATGT	GGAAGTTGGG	AATGTCATCA	AGTCCAAGCT	TAGGAGTTAT	CCGGCCCACT	9480
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTTAA	CATAGAAGAC	AAAGAGTCAA	9540
CGAGGAAGAT	CCGTGAACTC	CTCAAAAAGG	GGAATTCGCT	GTACTCCAAA	GTCAGTGATA	9600
AGGTTTTCCA	ATGCTTAAGG	GACACTAACT	CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAGTT	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAGTGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TGACATTTGA	ACTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGAC	TATTGATGCT	AGGTATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AACTGATAGA	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCCATGC	10080
TGGAGCCTCT	TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTAA	CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
ATGAAGGTAC	TTATCATGAG	TTAATTGAAG	CTCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGAC	AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGGC	TGCTGAAAAT	GTTAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGAT	GAAAGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAGG	CAGTTGGCCA	CCGCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAAGC	TTCAGGTGAA	GGGTTAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAAATCTT	10560
TTGCTGGAGT	GAAATTTGGC	TGCTTTATGC	CTCTTAGCCT	GGATAGTGAT	CTGACAATGT	10620

- 206 -

*						
ACCTAAAGG	A CAAGGCACT	GCTGCTCTC	CAAAGGGAATG	GGATTCAGT	TACCCGAAAG	10680
AGTTCCTGC	TTACGACCCT	r cccaagggai	A CCGGGTCACG	GAGGCTTGT	A GATGTTTTCC	10740
TTAATGATT	C GAGCTTTGAC	CCATATGAT	F TGATAATGTA	TGTTGTAAGI	GGAGCTTACC	10800
TCCATGACCC	C TGAGTTCAAC	CTGTCTTAC	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTTTT	TGCTAAAATG	ACTTACAAA	TGAGGGCATG	CCAAGTGATT	GCTGAAAATC	10920
TAATCTCAAA	CGGGATTGGC	: AAATATTT	AGGACAATGG	GATGGCCAAG	GATGAGCACG	10980
ATTTGACTAA	GGCACTCCAC	ACTCTAGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGGGG	GGGGCCAGTC	TTAAAAACCI	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100
GGAACGTGAG	AGCAGCAAAA	GGGTTTATAG	GGTTCCCTCA	AGTAATTCGG	CAGGACCAAG	11160
ACACTGATCA	TCCGGAGAAT	ATGGAAGCTT	' ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAGAA	GTACTGCCTT	AATTGGAGAT	ATGAGACCAT	CAGCTTGTTT	GCACAGAGGC	11280
TAAATGAGAT	TTACGGATTG	CCCTCATTTT	TCCAGTGGCT	GCATAAGAGG	CTTGAGACCT	11340
CTGTCCTGTA	TGTAAGTGAC	CCTCATTGCC	CCCCCGACCT	TGACGCCCAT	ATCCCGTTAT	11400
ATAAAGTCCC	CAATGATCAA	ATCTTCATTA	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTATCTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGGAT	TGCTTCGTTA	GTGCAAGGGG	ACAATCAGAC	CATAGCCGTA	ACAAAAAGGG	11580
TACCCAGCAC	ATGGCCCTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ATATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAA	ATTCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACAACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGGG	ATGAATTATC	12060
IGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCCTCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180

- 207 -

CACAACAAC	GGGGGACTC	TCATTCCTA	G ACTGGGCTAG	GACCCTTAC	CTCAGCAAATC	12240
TTGTATGTGT	CCAGAGCATO	CACTAGACTCO	TCAAGAACAT	AACTGCAAGG	TTTGTCCTGA	12300
TCCATAGTCC	AAACCCAATG	TTAAAAGGAT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGGGACTGGC	GGCATTCCTC	: ATGGACAGGC	ATATTATAGI	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATCA	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTG	GATACCACAA	12480
AAGGCCTGAT	TCGAGCCAGC	ATGAGGAAGG	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATTA	TGACTATGAA	CAATTCAGAG	CAGGGATGGT	GCTATTGACA	GGAAGAAAGA	12600
GAAATGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCGAGAGCT	CTAAGAAGCC	12660
ATATGTGGGC	GAGGCTAGCT	' CGAGGACGGC	: CTATTTACGG	CCTTGAGGTC	CCTGATGTAC	12720
TAGAATCTAT	GCGAGGCCAC	CTTATTCGGC	GTCATGAGAC	ATGTGTCATC	TGCGAGTGTG	12780
GATCAGTCAA	CTACGGATGG	TTTTTTGTCC	CCTCGGGTTG	CCAACTGGAT	GATATTGACA	12840
AGGAAACATC	ATCCTTGAGA	GTCCCATATA	TTGGTTCTAC	CACTGATGAG	AGAACAGACA	12900
TGAAGCTTGC	CTTCGTAAGA	GCCCCAAGTC	GATCCTTGCG	ATCTGCTGTT	AGAATAGCAA	12960
CAGTGTACTC	ATGGGCTTAC	GGTGATGATG	ATAGCTCTTG	GAACGAAGCC	TGGTTGTTGG	13020
CTAGGCAAAG	GGCCAATGTG	AGCCTGGAGG	AGCTAAGGGT	GATCACTCCC	ATCTCAACTT	13080
CGACTAATTT	AGCGCATAGG	TTGAGGGATC	GTAGCACTCA	AGTGAAATAC	TCAGGTACAT	13140
CCCTTGTCCG	AGTGGCGAGG	TATACCACAA	TCTCCAACGA	CAATCTCTCA	TTTGTCATAT	13200
CAGATAAGAA	GGTTGATACT	AACTTTATAT	ACCAACAAGG	AATGCTTCTA	GGGTTGGGTG	13260
TTTTAGAAAC	ATTGTTTCGA	CTCGAGAAAG	ATACCGGATC	ATCTAACACG	GTATTACATC	13320
			CGATGATAGA			13380
			GTACCAACCC			13440
CTTTAATTGA	CAGAGATACA	ACAAGGCTAT	ACACCCAGAG	CCATAGGAGG	CACCTTGTGG	13500
					ACAGCACTAT	13560
					TCAGCTCTCA	13620
			CTGAGTTTCT			13680
TCACTATCTA	CTTGGGCCAG	TGTGCGGCCA	TCAATTGGGC	ATTTGATGTA	CATTATCATA	13740

- 208 -

GACCATCAGG	GAAATATCAG	ATGGGTGAGC	TGTTGTCATC	GTTCCTTTCT	AGAATGAGCA	13800
AAGGAGTGTT	TAAGGTGCTT	GTCAATGCTC	TAAGCCACCC	AAAGATCTAC	AAGAAATTCT	13860
GGCATTGTGG	TATTATAGAG	CCTATCCATG	GTCCTTCACT	TGATGCTCAA	AACTTGCACA	13920
CAACTGTGTG	CAACATGGTT	TACACATGCT	ATATGACCTA	CCTCGACCTG	TTGTTGAATG	13980
aagagttaga	AGAGTTCACA	TTTCTCTTGT	GTGAAAGCGA	CGAGGATGTA	GTACCGGACA	14040
GATTCGACAA	CATCCAGGCA	AAACACTTAT	GTGTTCTGGC	AGATTTGTAC	TGTCAACCAG	14100
GGACCTGCCC	ACCAATTCGA	GGTCTAAGAC	CGGTAGAGAA	ATGTGCAGTT	CTAACCGACC	14160
ATATCAAGGC	AGAGGCTAGG	TTATCTCCAG	CAGGATCTTC	GTGGAACATA	AATCCAATTA	14220
TTGTAGACCA	TTACTCATGC	TCTCTGACTT	ATCTCCGGCG	AGGATCGATC	AAACAGATAA	14280
GATTGAGAGT	TGATCCAGGA	TTCATTTTCG	ACGCCCTCGC	TGAGGTAAAT	GTCAGTCAGC	14340
CAAAGATCGG	CAGCAACAAC	ATCTCAAATA	TGAGCATCAA	GGCTTTCAGA	CCCCCACACG	14400
ATGATGTTGC	AAAATTGCTC	AAAGATATCA	ACACAAGCAA	GCACAATCTT	CCCATTTCAG	14460
GGGGCAATCT	CGCCAATTAT	GAAATCCATG	CTTTCCGCAG	AATCGGGTTG	AACTCATCTG	14520
CTTGCTACAA	AGCTGTTGAG	ATATCAACAT	TAATTAGGAG	ATGCCTTGAG	CCAGGGGAGG	14580
ACGGCTTGTT	CTTGGGTGAG	GGATCGGGTT	CTATGTTGAT	CACTTATAAG	GAGATACTTA	14640
AACTAAACAA	GTGCTTCTAT	AATAGTGGGG	TTTCCGCCAA	TTCTAGATCT	GGTCAAAGGG	14700
AATTAGCACC	CTATCCCTCC	GAAGTTGGCC	TTGTCGAACA	CAGAATGGGA	GTAGGTAATA	14760
TTGTCAAAGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	GTAGATTGCT	14820
	AGTTAGTAAT					14880
	TAACAAAGAT					14940
TGGCTCTGCT	CCTGGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
					GTGAACCTTG	15060
					ACAGATCTCA	15120
					TCATCTGTGA	15180
					TGCATACAAG	15240
CAATTGTGGG	AGACGCAGTT	AGTAGAGGTG	ATATCAATCC	TACTCTGAAA	AAACTTACAC	15300

- 209 -

CTATAGAGCA GGTGCTGATC AATTGCGGGT TGGCAATTAA CGGACCTAAG CTGTGCAAAG 15360 AATTGATCCA CCATGATGTT GCCTCAGGGC AAGATGGATT GCTTAATTCT ATACTCATCC 15420 TCTACAGGGA GTTGGCAAGA TTCAAAGACA ACCAAAGAAG TCAACAAGGG ATGTTCCACG 15480 CTTACCCCGT ATTGGTAAGT AGCAGGCAAC GAGAACTTAT ATCTAGGATC ACCCGCAAAT 15540 TTTGGGGGCA CATTCTTCTT TACTCCGGGA ACAGAAGTT GATAAATAAG TTTATCCAGA 15600 ATCTCAAGTC CGGCTATCTG ATACTAGACT TACACCAGAA TATCTTCGTT AAGAATCTAT 15660 CCAAGTCAGA GAAACAGATT ATTATGACGG GGGGTTTGAA ACGTGAGTGG GTTTTTAAGG 15720 TAACAGTCAA GGAGACCAAA GAATGGTATA AGTTAGTCGG ATACAGTGCC CTGATTAAGG 15780 ACTANTIGGT TGANCTCCGG ANCCCTANTC CTGCCCTAGG TGGTTAGGCA TTATTTGCAA 15840 TAGATTAAAG AAAACTTTGA AAATACGAAG TTTCTATTCC CAGCTTTGTC TGGT 15894

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2183 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
- Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15
- Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala
  20 25 30
- Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45
- Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 55
- Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80
- Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn

- 210 -

					85					90					95	,
	Ile	Glu	Asp	Lys 100	Glu	Ser	Thr	Arg	Lys 105		Arg	Glu	Leu	Leu 110	Lys	Ly
	Gly	Asn	Ser 115	Leu	Tyr	Ser	Lys	Val 120	Ser	Asp	Lys	Val	Phe 125		Сув	Le
	Arg	Asp 130	Thr	Asn	Ser	Arg	Leu 135	Gly	Leu	Gly	Ser	Glu 140	Leu	Arg	Glu	As
	Ile 145	Lys	Glu	Lys	Val	Ile 150	Asn	Leu	Gly	Val	Tyr 155	Met	His	Ser	Ser	G1:
	Trp	Phe	Glu	Pro	Phe 165	Leu	Phe	Trp	Phe	Thr 170	Val	Lys	Thr	Glu	Met 175	Ar
	Ser	Val	Ile	Lys 180	Ser	Gln	Thr	His	Thr 185	Cys	His	Arg	Arg	Arg 190	His	Thi
	Pro	Val	Phe 195	Phe	Thr	Gly	Ser	Ser 200	Val	Glu	Leu	Leu	Ile 205	Ser	Arg	λвј
		210				Ser	215					220				
	Phe 225	Glu	Leu	Val	Leu	Met 230	Tyr	Сув	Asp	Val	11e 235	Glu	Gly	Arg	Leu	Met 240
					245	Thr				250					255	_
1	Arg	Val	Arg	Tyr 260	Met	Trp	Lys	Leu	Ile 265	Asp	Gly	Phe	Phe	Pro 270	Ala	Leu
•	Gly	Asn	Pro 275	Thr	Tyr	Gln	Ile	Val 280	Ala	Met	Leu	Glu	Pro 285	Leu	Ser	Lev
		290				Arg	295					300				
:	Leu 305	Asn	His	Сув	Phe	Thr 310	Glu	Ile	His	Asp	Val 315	Leu	qaA	Gln	Asn	Gly 320
1	Phe	Ser	Asp	Glu	Gly 325	Thr	Tyr	His	Glu	J30	Ile	Glu	Ala	Leu	Авр 335	Tyr
•	Cle	Phe	Ile	Thr 340	Авр	Asp	Ile	His	Leu 345	Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
3	?he	Arg	Ser 355	Phe	Gly	His	Pro	Arg 360	Leu	Glu	Ala	Val	Thr 365	Ala	Ala	Glu

- 211 -

Asn	Val 370	Arg	Lys	Tyr	Met	<b>As</b> n 375	Gln	Pro	Lys	Val	Ile 380	Val	Tyr	Glu	Thr
Leu 385	Met	Lys	Gly	His	Ala 390	Ile	Phe	Сув	Gly	11e 395	Ile	Ile	Asn	Gly	Tyr 400
Arg	Asp	Arg	His	Gly 405	Gly	Ser	Trp	Pro	Pro 410	Leu	Thr	Leu	Pro	Leu 415	His
Ala	Ala	Asp	Thr 420	Ile	Arg	Asn	Ala	Gln 425	Ala	Ser	Gly	Glu	Gly 430	Leu	Thr
His	Glu	Gln 435	Сув	Val	Asp	Asn	Trp 440	Lys	Ser	Phe	Ala	Gly 445	Val	Lys	Phe
Gly	Сув 450	Phe	Met	Pro	Leu	Ser 455	Leu	Asp	Ser	Asp	Leu 460	Thr	Met	Tyr	Leu
Lys 465	Авр	Lys	Ala	Leu	Ala 470	Ala	Leu	Gln	Arg	Glu 475	Trp	Asp	Ser	Val	Tyr 480
Pro	Lys	Glu	Phe	Leu 485	Arg	Tyr	Авр	Pro	Pro 490	Lys	Gly	Thr	Gly	Ser 495	Arg
Arg	Leu	Val	Asp 500	Val	Phe	Leu	Asn	Дар 505	Ser	Ser	Phe	Авр	Pro 510	Tyr	Хвр
Val	Ile	Met 515	Tyr	Val	Val	Ser	Gly 520	Ala	Tyr	Leu	His	Авр 525	Pro	Glu	Phe
Asn	Leu 530	Ser	Tyr	Ser	Leu	<b>Lys</b> 535	Glu	Lys	Glu	Ile	Lys 540	Glu	Thr	Gly	Arg
Leu 545	Phe	Ala	Lys	Met	Thr 550	Tyr	Lys	Met	Arg	Ala 555	Сув	Gln	Val	Ile	Ala 560
Glu	Asn	Leu	Ile	Ser 565	Asn	Gly	Ile	Gly	Lys 570	Tyr	Phe	Lys	yab	Asn 575	Gly
Met	Ala	Lys	<b>Аар</b> 580	Glu	His	Yab	Leu	Thr 585	Lys	Ala	Leu	His	Thr 590	Leu	Ala
Val	Ser	Gly 595		Pro	Lys	Авр	Leu 600	Lys	Glu	Ser	His	Arg 605	Gly	Gly	Pro
Val	Leu 610	Lys	Thr	Tyr	Ser	Arg 615	Ser	Pro	Val	His	Thr 620	Ser	Thr	Arg	Asn
Val 625	Arg	Ala	Ala	Lys	Gly 630	Phe	Ile	Gly	Phe	Pro 635	Gln	Val	Ile	Arg	Gln 640

- 212 -

Увр	Gln	Авр	Thr	Авр 645	His	Pro	Glu	Asn	Met 650		Ala	Tyr	Glu	Thr 655	Val
Ser	Ala	Phe	Ile 660	Thr	Thr	Авр	Leu	Lys 665		Tyr	Сув	Leu	<b>Asn</b> 670	Trp	Arg
Tyr	Glu	Thr 675	Ile	Ser	Leu	Phe	Ala 680	Gln	Arg	Leu	Asn	Glu 685		Tyr	Gly
Leu	Pro 690	Ser	Phe	Phe	Gln	Trp 695	Leu	His	Lys	Arg	Leu 700	Glu	Thr	Ser	Val
Leu 705	Tyr	Val	Ser	Авр	Pro 710	His	Сув	Pro	Pro	Авр 715	Leu	Авр	Ala	His	Ile 720
Pro	Leu	Tyr	Lys	Val 725	Pro	Asn	Авр	Gln	Ile 730	Phe	Ile	Lys	Tyr	Pro 735	Met
			740		Tyr			745					750		
Pro	Tyr	Leu 755	Tyr	Leu	Ala	Ala	Tyr 760	Glu	Ser	Gly	Val	Arg 765	Ile	Ala	Ser
Leu	Val 770	Gln	Gly	Asp	Asn	Gln 775	Thr	Ile	Ala	Val	Thr 780	Lys	Arg	Val	Pro
785					Asn 790			•		795					800
				805	Ile				810					815	
His	Leu	Lys	Ala 820	Asn	Glu	Thr	Ile	Val 825	Ser	Ser	His	Phe	Phe 830	Val	Tyr
		835			Tyr		840					845			
Ser	11e 850	Ala	Arg	Сув	Val	Phe 855	Trp	Ser	Glu	Thr	Ile 860	Val	Ąap	Glu	Thr
Arg 865	Ala	Ala	Сув	Ser	Asn 870	Ile	Ala	Thr	Thr	Met 875	Ala	Lys	Ser	Ile	Glu 880
Arg	Gly	Tyr	Asp	Arg 885	Tyr	Leu	Ala	Tyr	Ser 890	Leu	Asn	Val	Leu	Lys 895	Val
			900		Ile			905					910		
Thr	Arg	qaA	Val	Val	Ile	Pro	Leu	Leu	Thr	Asn	Asn	Asp	Leu	Leu	Ile

- 213 -

		915					920					925	;		
Arg	Met 930		Leu	Leu	Pro	Ala 935		Ile	Gly	Gly	Met 940		Tyr	Leu	Asr
Met 945	Ser	Arg	Leu	Phe	Val 950	Arg	Asn	Ile	Gly	Asp 955		Val	Thr	Ser	960
Ile	Ala	Asp	Leu	<b>Lys</b> 965	Arg	Met	Ile	Leu	Ala 970		Leu	Met	Pro	Glu 975	
Thr	Leu	His	Gln 980	Val	Met	Thr	Gln	Gln 985	Pro	Gly	Asp	Ser	Ser 990	Phe	Leu
Авр	Trp	Ala 995	Ser	Asp	Pro	Tyr	Ser 100		Asn	Leu	Val	Сув 100		Gln	Ser
Ile	Thr 101		Leu	Leu	Lys	Asn 101		Thr	Ala	Arg	Phe 102		Leu	Ile	His
Ser 102		Asn	Pro	Met	Leu 1030		Gly	Leu	Phe	His 103		Asp	Ser	Lys	Glu 104
Glu	qaA	Glu	Gly	Leu 104!		Ala	Phe	Leu	Met 105		Arg	His	Ile	11e 105	
Pro	Arg	Ala	Ala 1060		Glu	Ile	Leu	Asp 106		Ser	Val	Thr	Gly 1070		Arg
Glu	Ser	Ile 107	Ala 5	Gly	Met	Leu	Asp 1080		Thr	Lys	Gly	Leu 108		Arg	Ala
Ser	Met 1090		Lys	Gly	Gly	Leu 109!		Ser	Arg	Val	11e		Arg	Leu	Ser
Asn 1105		Asp	Tyr	Glu	Gln 1110		Arg	Ala	Gly	Met 1115		Leu	Leu	Thr	Gly 112
Arg	Lys	Arg	Asn	Val 1125		Ile	Asp	Lys	Glu 113(		Cys	Ser	Val	Gln 1135	
Ala	Arg	Ala	Leu 1140		Ser	Kis	Met	Trp 1145		Arg	Leu	Ala	Arg 1150	_	Arg
Pro	Ile	Tyr 1155	Gly	Leu	Glu	Val	Pro 1160		Val	Leu	Glu	Ser 1165		Arg	Gly
His	Leu 1170	Ile	Arg	Arg		Glu 1175		Сув	Val	Ile	Сув 1180		Сув	Gly	Ser
Val 1185		Tyr	Gly	Trp	Phe 1190		Val	Pro	Ser	Gly 1195		Gln	Leu	Asp	Asp 1200

- 214 -

Ile	Авр	Lys	Glu	Thr 120		Ser	Leu	Arg	Val 121		Tyr	Ile	Gly	Ser 121	
Thr	Asp	Glu	Arg 1220		Asp	Met	Lys	Leu 122		Phe	Val	Arg	Ala 123		Ser
Arg	Ser	Leu 123		Ser	Ala	Val	Arg 124(		Ala	Thr	Val	Tyr 124		Trp	Ala
Tyr	Gly 125		Asp	qaA	Ser	Ser 125		Asn	Glu	Ala	Trp 126		Leu	Ala	Arg
Gln 126	Arg 5	Ala	Asn	Val	Ser 1270		Glu	Glu	Leu	Arg 1275		Ile	Thr	Pro	Ile 1280
Ser	Thr	Ser	Thr	Asn 1285		Ala	His	Arg	Leu 1290		Asp	Arg	Ser	Thr 1295	
Val	Lys	Tyr	Ser 1300		Thr	Ser	Leu	Val 130		Val	Ala	Arg	Tyr 1310		Thr
Ile	Ser	Asn 131		Asn	Leu	Ser	Phe 1320		Ile	Ser	Asp	Lys 1325	-	Val	Авр
Thr	Asn 133		Ile	Tyr	Gln	Gln 133		Met	Leu	Leu	Gly 1340		Gly	Val	Leu
Glu 134!	Thr 5	Leu	Phe	Arg	Leu 1350		Lys	Asp	Thr	Gly 1355		Ser	Asn	Thr	Val 1360
Leu	His	Leu	His	Val 1365		Thr	Asp	Сув	Cys 137(		Ile	Pro	Met	Ile 1375	_
His	Pro	Arg	Ile 1380		Ser	Ser	Arg	Lys 138		Glu	Leu	Arg	Ala 1390		Leu
Сув	Thr	Asn 1395		Leu	Ile	Tyr	Asp 1400		Ala	Pro	Leu	Ile 1405	_	Arg	Asp
Thr	Thr 1410		Leu	Tyr	Thr	Gln 1415		His	Arg	Arg	His 1420		Val	Glu	Phe
Val 1425	Thr	Trp	Ser	Thr	Pro 1430		Leu	Tyr	His	Ile 1435		Ala	Lys	Ser	Thr 1440
Ala	Leu	Ser	Met	Ile 1445		Leu	Val	Thr	Lys 1450		Glu	Lys	Asp	His 1455	
_		TIA	0.~		T.011	Tla	Gly	Agn	Asp	Aen	Tla	200	00=	Dh.	<b>T1</b> -

1465

1470

1460

- 215 -

- Thr Glu Phe Leu Leu Ile Glu Pro Arg Leu Phe Thr Ile Tyr Leu Gly 1475 1480 1485
- Gln Cys Ala Ala Ile Asn Trp Ala Phe Asp Val His Tyr His Arg Pro 1490 1495 1500
- Ser Gly Lys Tyr Gln Met Gly Glu Leu Leu Ser Ser Phe Leu Ser Arg 1505 1510 1515 1520
- Met Ser Lys Gly Val Phe Lys Val Leu Val Asn Ala Leu Ser His Pro 1525 1530 1535
- Lys Ile Tyr Lys Lys Phe Trp His Cys Gly Ile Ile Glu Pro Ile His 1540 1545 1550
- Gly Pro Ser Leu Asp Ala Gln Asn Leu His Thr Thr Val Cys Asn Met 1555 1560 1565
- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Thr Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635 1640 1645
- Arg Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1660
- Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys 1665 1670 1675 1680
- Gln Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala 1685 1690 1695
- Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn 1700 1705 1710
- Met Ser Ile Lys Ala Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu 1715 1720 1725
- Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly 1730 1740
- Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn

- 216 -

1755

1760

1750

1745

Ser	Ser	Ala	Сув	Tyr 176!	Lys 5	Ala	Val	Glu	Ile 1770		Thr	Leu	Ile	Arg 1779	_
Сув	Leu	Glu	Pro 1780		Glu	Asp	Gly	Leu 178!		Leu	Gly	Glu	Gly 1790		Gly
Ser	Met	Leu 179		Thr	Tyr	Lys	Glu 1800		Leu	Lys	Leu	Asn 1805	-	Сув	Phe
Tyr	Asn 181		Gly	Val	Ser	Ala 1815		Ser	Arg	Ser	Gly 1820		Arg	Glu	Leu
Ala 182		Tyr	Pro	Ser	Glu 1830		Gly	Leu	Val	Glu 183		Arg	Met	Gly	Val 1840
Gly	Asn	Ile	Val	Lys 1845	Val	Leu	Phe	Asn	Gly 1850		Pro	Glu	Val	Thr 1855	_
Val	Gly	Ser	Val 1860		Сув	Phe	Asn	Phe 1865		Val	Ser	Asn	Ile 1870		Thr
Ser	Ser	Val 1875		Phe	Ile	His	Ser 1880		Ile	Glu	Thr	Leu 1885		Asn	Lys
Хвр	Thr 1890		Glu	Lys	Leu			Leu	Ala	Ala			Ser	Met	Ala
	1890	,				1895	5				1900	)			
Leu 190	Leu		Gly	Lys	Ile 1910	Gly		Ile	Leu	Val 1915	Ile		Leu	Met	Pro 1920
190	Leu 5	Leu			1910 Val	Gly )	Ser			1915 Ser	Ile	Lys			1920 His
190! Phe	Leu 5 Ser	Leu	yab	Phe 1925 Asn	1910 Val	Gly ) Gln	Ser	Phe	Ile 1930 Arg	1915 Ser	Ile Tyr	Lys Val	Gly	Ser 1935	1920 His
Phe	Leu Ser Arg	Leu Gly Glu	Asp Val 1940	Phe 1925 Asn	1910 Val	Gly Gln Val	Ser Gly Tyr	Phe Pro 1945	Ile 193( Arg	1915 Ser ) Tyr	Ile Tyr Ser	Lys Val Asn	Gly Phe 1950 Arg	Ser 1935 Ile	1920 His Ser
Phe Tyr	Leu Ser Arg	Gly Glu Ser 1955	Val 194( Tyr	Phe 1925 Asn )	Val Leu	Gly Gln Val	Gly Tyr Thr 1960	Phe Pro 1945 Asp	Ile 193( Arg Leu	Ser Tyr	Ile Tyr Ser	Lys Val Asn Asn 1965	Gly Phe 1950 Arg	Ser 1935 Ile )	1920 His Ser Met
Phe Tyr Thr	Leu Ser Arg Glu Pro 1970	Gly Glu Ser 1955	Val 1940 Tyr	Phe 1925 Asn Leu	Val ; Leu Val	Gly Gln Val Met Gln 1975	Gly Tyr Thr 1960	Phe Pro 1945 Asp	Ile 193( Arg ; Leu	Ser Tyr Lys	Ile Tyr Ser Ala Ser 1980	Lys Val Asn Asn 1965	Phe 1950 Arg	Ser 1935 Ile Leu Arg	1920 His Ser Met
Tyr Thr Asn Ser 1985	Ser Arg Glu Pro 1970	Gly Glu Ser 1955 Glu Gly	Val 1940 Tyr Lys	Phe 1925 Asn Leu Ile	Val Leu Val Lys Gly 1990	Gly Gln Val Met Gln 1975	Gly Tyr Thr 1960 Gln	Pro 1945 Asp Ile	Ile 1930 Arg Leu Ile	Ser Tyr Lys Glu 11e 1995	Tyr Ser Ala Ser 1980	Lys Val Asn 1965 Ser	Phe 1950 Arg Val	Ser 1935 Ile Leu Arg	1920 His Ser Met Thr Cys 2000

- 217 -

	Leu	Ala	Ile 2035		Gly	Pro	Lys	Leu 204		Lys	Glu	Leu	Ile 204!	His 5	His	Asp	
	Val	Ala 2050		Gly	Gln	Asp	Gly 205		Leu	Asn	Ser	11e 2060		Ile	Leu	Tyr	
	Arg 2065		Leu	Ala	Arg	Phe 2070		Asp	Asn	Gln	Arg 2075		Gln	Gln	Gly	Met 2080	
	Phe	His	Ala	Tyr	Pro 208		Leu	Val	Ser	Ser 2090		Gln	Arg	Glu	Leu 2095		
	Ser	Arg	Ile	Thr 2100		Lys	Phe	Trp	Gly 2105		Ile	Leu	Leu	Tyr 2110		Gly	
	Asn	Arg	Lys 2115		Ile	Asn	Lys	Phe 2120		Gln	Asn	Leu	Lys 2125	Ser	Gly	Tyr	
	Leu	11e 2130		Asp	Leu	His	Gln 2135		Ile	Phe	Val	Lys 2140		Leu	Ser	Lys	
	Ser 2145		Lys	Gln	Ile	11e 2150		Thr	Gly	Gly	Leu 2155	_	Arg	Glu	Trp	Val 2160	
	Phe	Lys	Val	Thr	Val 2165		Glu	Thr	Lys	Glu 2170		Tyr	Lys	Leu	Val 2175	_	
	Tyr	Ser	Ala	Leu 2180		Lys	Asp										
(2) 1	NFOR	MATI	ON F	or s	EQ I	D NO	:15:										
	(i)		LEN	IGTH :	158	94 b	ase		<b>.</b> 8								
		(C)	STR	ANDE	DNES	ic a S: a inea	ingl	.e									
(	( <b>ii</b> )	MOLE	CULE	TYP	E: R	LNA (	geno	mic)									
(	xi)	SEQU	ENCE	DES	CRIP	TION	i: SE	g ID	NO:	15:							
ACCAA	ACAA	A GI	TGGG	TAAG	GAT	'AGTT	CAA	TCAA	TGAT	CA T	CTTC	TAGT	G CA	.CTTA	GGAT		60
TCAAG	ATCC	T AT	TATC	aggg	ACA	AGAG	CAG	GATT	AGGG	AT A	TCCG	AGAT	G GC	CACA	CTTT		120
TAAGG	AGCT	T AG	CATT	GTTC	AAA	agaa	ACA	AGGA	CAAA	CC A	CCCA	TTAC	A TC	AGGA	TCCG		180

- 218 -

GTGGAGCCAT	CAGAGGAATC	AAACACATTA	TTATAGTACC	AATCCCTGGA	GATTCCTCAA	240
TTACCACTCG	ATCCAGACTT	CTGGACCGGT	TGGTCAGGTT	AATTGGAAAC	CCGGATGTGA	300
GCGGGCCCAA	ACTAACAGGG	GCACTAATAG	GTATATTATC	CTTATTTGTG	GAGTCTCCAG	360
GTCAATTGAT	TCAGAGGATC	ACCGATGACC	CTGACGTTAG	CATAAGGCTG	TTAGAGGTTG	420
TCCAGAGTGA	CCAGTCACAA	TCTGGCCTTA	CCTTCGCATC	AAGAGGTACC	AACATGGAGG	480
ATGAGGCGGA	CAAATACTTT	TCACATGATG	ATCCAATTAG	TAGTGATCAA	TCCAGGTTCG	540
GATGGTTCGA	GAACAAGGAA	ATCTCAGATA	TTGAAGTGCA	AGACCCTGAG	GGATTCAACA	600
TGATTCTGGG	TACCATCCTA	GCCCAAATTT	GGGTCTTGCT	CGCAAAGGCG	GTTACGGCCC	660
CAGACACGGC	AGCTGATTCG	GAGCTAAGAA	GGTGGATAAA	GTACACCCAA	CAAAGAAGGG	720
TAGTTGGTGA	ATTTAGATTG	GAGAGAAAT	GGTTGGATGT	GGTGAGGAAC	AGGATTGCCG	780
AGGACCTCTC	CTTACGCCGA	TTCATGGTCG	CTCTAATCCT	GGATATCAAG	AGAACACCCG	840
GAAACAAACC	CAGGATTGCT	GAAATGATAT	GTGACATTGA	TACATATATC	GTAGAGGCAG	900
GATTAGCCAG	TTTTATCCTG	ACTATTAAGT	TTGGGATAGA	AACTATGTAT	CCTGCTCTTG	960
GACTGCATGA	ATTTGCTGGT	GAGTTATCCA	CACTTGAGTC	CTTGATGAAC	CTTTACCAGC	1020
AAATGGGGGA	AACTGCACCC	TACATGGTAA	TCCTGGAGAA	CTCAATTCAG	AACAAGTTCA	1080
GTGCAGGATC	ATACCCTCTG	CTCTGGAGCT	ATGCCATGGG	AGTAGGAGTG	GAACTTGAAA	1140
ACTCCATGGG	AGGTTTGAAC	TTTGGCCGAT	CTTACTTTGA	TCCAGCATAT	TTTAGATTAG	1200
GGCAAGAGAT	GGTAAGGAGG	TCAGCTGGAA	AGGTCAGTTC	CACATTGGCA	TCTGAACTCG	1260
GTATCACTGC	CGAGGATGCA	AGGCTTGTTT	CAGAGATTGC	AATGCATACT	ACTGAGGACA	1320
AGATCAGTAG	AGCGGTTGGA	CCCAGACAAG	CCCAAGTATC	ATTTCTACAC	GGTGATCAAA	1380
GTGAGAATGA	GCTACCGAGA	TTGGGGGGCA	AGGAAGATAG	GAGGGTCAAA	CAGAGTCGAG	1440
GAGAAGCCAG	GGAGAGCTAC	AGAGAAACCG	GGCCCAGCAG	AGCAAGTGAT	GCGAGAGCTG	1500
CCCATCTTCC	AACCGGCACA	CCCCTAGACA	TTGACACTGC	ATCGGAGTCC	AGCCAAGATC	1560
CGCAGGACAG	TCGAAGGTCA	GCTGACGCCC	TGCTTAGGCT	GCAAGCCATG	GCAGGAATCT	1620
CGGAAGAACA	AGGCTCAGAC	ACGGACACCC	CTATAGTGTA	CAATGACAGA	AATCTTCTAG	1680
ACTAGGTGCG	AGAGGCCGAG	GACCAGAACA	ACATCCGCCT	ACCCTCCATC	ATTGTTATAA	1740

- 219 -

AAAACTTAGG	AACCAGGTCC	ACACAGCCGC	CAGCCCATCA	ACCATCCACT	CCCACGATTG	1800
GAGCCGATGG	CAGAAGAGCA	GGCACGCCAT	GTCAAAAACG	GACTGGAATG	CATCCGGGCT	1860
CTCAAGGCCG	AGCCCATCGG	CTCACTGGCC	ATCGAGGAAG	CTATGGCAGC	ATGGTCAGAA	1920
ATATCAGACA	ACCCAGGACA	GGAGCGAGCC	ACCTGCAGGG	AAGAGAAGGC	AGGCAGTTCG	1980
GGTCTCAGCA	AACCATGCCT	CTCAGCAATT	GGATCAACTG	AAGGCGGTGC	ACCTCGCATC	2040
CGCGGTCAGG	GACCTGGAGA	GAGCGATGAC	GACGCTGAAA	CTTTGGGAAT	CCCCCAAGA	2100
AATCTCCAGG	CATCAAGCAC	TGGGTTACAG	TGTTATTATG	TTTATGATCA	CAGCGGTGAA	2160
GCGGTTAAGG	GAATCCAAGA	TGCTGACTCT	ATCATGGTTC	AATCAGGCCT	TGATGGTGAT	2220
AGCACCCTAT	CAGGAGGAGA	CAATGAATCT	GAAAACAGCG	ATGTGGATAT	TGGCGAACCT	2280
GATACCGAGG	GATATGCTAT	CACTGACCGG	GGATCTGCTC	CCATCTCTAT	GGGGTTCAGG	2340
GCTTCTGATG	TTGAAACTGC	AGAAGGAGGG	GAGATCCACG	AGCTCCTGAG	ACTCCAATCC	2400
AGAGGCAACA	ACTTTCCGAA	GCTTGGGAAA	ACTCTCAATG	TTCCTCCGCC	CCCGGACCCC	2460
GGTAGGGCCA	GCACTTCCGG	GACACCCATT	AAAAAGGGCA	CAGACGCGAG	ATTAGCCTCA	2520
TTTGGAACGG	AGATCGCGTC	TTTATTGACA	GGTGGTGCAA	CCCAATGTGC	TCGAAAGTCA	2580
CCCTCGGAAC	CATCAGGGCC	AGGTGCACCT	GCGGGGAATG	TCCCCGAGTA	TGTGAGCAAT	2640
GCCGCACTGA	TACAGGAGTG	GACACCCGAA	TCTGGTACCA	CAATCTCCCC	GAGATCCCAG	2700
AATAATGAAG	AAGGGGGAGA	CTATTATGAT	GATGAGCTGT	TCTCTGATGT	CCAAGATATT	2760
AAAACAGCCT	TGGCCAAAAT	ACACGAGGAT	AATCAGAAGA	TAATCTCCAA	GCTAGAATCA	2820
CTGCTGTTAT	TGAAGGGAGA	AGTTGAGTCA	ATTAAGAAGC	AGATCAACAG	GCAAAATATC	2880
AGCATATCCA	CCCTGGAAGG	ACACCTCTCA	AGCATCATGA	TCGCCATTCC	TGGACTTGGG	2940
AAGGATCCCA	ACGACCCCAC	TGCAGATGTC	GAAATCAATC	CCGACTTGAA	ACCCATCATA	3000
GGCAGAGATT	CAGGCCGAGC	ACTGGCCGAA	GTTCTCAAGA	AACCCGTTGC	CAGCCGACAA	3060
CTCCAAGGAA	TGACAAATGG	ACGGACCAGT	TCCAGAGGAC	AGCTGCTGAA	GGAATTTCAG	3120
CCAAAGCCGA	TCGGGAAAA	GATGAGCTCA	GCCGTCGGGT	TTGTTCCTGA	CACCGGCCCT	3180
GCATCACGCA	GTGTAATCCG	CTCCATTATA	AAATCCAGCC	GGCTAGAGGA	GGATCGGAAG	3240
CGTTACCTGA	TGACTCTCCT	TGATGATATC	AAAGGAGCCA	ATGATCTTGC	CAAGTTCCAC	3300

- 220 -

CAGATGCTGA TGAAGATAAT AATGAAGTAG CTACAGCTCA ACTTACCTGC CAACCCCATG	2260
CCAGTCGACC CAACTAGTAC AACCTAAATC CATTATAAAA AACTTAGGAG CAAAGTGATT	3360
GCCTCCCAAG TTCCACAATG ACAGAGATCT ACGACTTCGA CAAGTCGGCA TGGGACATCA	3420
	3480
AAGGGTCGAT CGCTCCGATA CAACCGACCA CCTACAGTGA TGGCAGGCTG GTGCCCCAGG	3540
TCAGAGTCAT AGATCCTGGT CTAGGCGACA GGAAGGATGA ATGCTTTATG TACATGTCTC	3600
TGCTGGGGGT TGTTGAGGAC AGCGATCCCC TAGGGCCTCC AATCGGGCGA GCATTTGGGT	3660
CCCTGCCCTT AGGTGTTGGC AGATCCACAG CAAAGCCCGA AAAACTCCTC AAAGAGGCCA	3720
CTGAGCTTGA CATAGTTGTT AGACGTACAG CAGGGCTCAA TGAAAAACTG GTGTTCTACA	3780
ACAACACCCC ACTAACTCTC CTCACACCTT GGAGAAAGGT CCTAACAACA GGGAGTGTCT	3840
TCAACGCAAA CCAAGTGTGC AATGCGGTTA ATCTGATACC GCTCGATACC CCGCAGAGGT	3900
TCCGTGTTGT TTATATGAGC ATCACCCGTC TTTCGGATAA CGGGTATTAC ACCGTTCCTA	3960
GAAGAATGCT GGAATTCAGA TCGGTCAATG CAGTGGCCTT CAACCTGCTG GTGACCCTTA	4020
GGATTGACAA GGCGATAGGC CCTGGGAAGA TCATCGACAA TACAGAGCAA CTTCCTGAGG	4080
CAACATTTAT GGTCCACATC GGGAACTTCA GGAGAAAGAA GAGTGAAGTC TACTCTGCCG	4140
ATTATTGCAA AATGAAAATC GAAAAGATGG GCCTGGTTTT TGCACTTGGT GGGATAGGGG	4200
GCACCAGTCT TCACATTAGA AGCACAGGCA AAATGAGCAA GACTCTCCAT GCACAACTCG	4260
GGTTCAAGAA GACCTTATGT TACCCGCTGA TGGATATCAA TGAAGACCTT AATCGATTAC	4320
TCTGGAGGAG CAGATGCAAG ATAGTAAGAA TCCAGGCAGT TTTGCAGCCA TCAGTTCCTC	4380
AAGAATTCCG CATTTACGAC GACGTGATCA TAAATGATGA CCAAGGACTA TTCAAAGTTC	4440
TGTAGACCGT AGTGCCCAGC AATGCCCGAA AACGACCCCC CTCACAATGA CAGCCAGAAG	4500
GCCCGGACAA AAAAGCCCCC TCCGAAAGAC TCCACTGACC AAGCGAGAGG CCAGCCAGCA	4560
GCCGACGGCA AGCACGAACA CCAGGCGGCC CCAGCACAGA ACAGCCCTGA TACAAGGCCA	4620
CCACCAGCCA CCCCAATCTG CATCCTCCTC GTGGGACCCC CGAGGACCAA CCCCCAAGGC	
TGCCCCCGAT CCAAACCACC AACCGCATCC CCACCACCCC CGGGAAAGAA ACCCCCAGCA	4680
ATTGGAAGGC CCCTCCCCT CTTCCTCAAC ACAAGAACTC CACAACCGAA CCGCACAAGC	4740
	4800
GACCGAGGTG ACCCAACCGC AGGCATCCGA CTCCCTAGAC AGATCCTCTC TCCCCGGCAA	4860

- 221 -

ACTAAACAAA	ACTTAGGGCC	AAGGAACATA	CACACCCAAC	AGAACCCAGA	CCCCGGCCCA	4920
CGGCGCCGCG	CCCCCAACCC	CCGACAACCA	GAGGGAGCCC	CCAACCAATO	CCGCCGGCTC	4980
CCCCGGTGCC	CACAGGCAGG	GACACCAACC	CCCGAACAGA	CCCAGCACCT	AACCATCGAC	5040
AATCCAAGAC	GGGGGGGCCC	CCCCAAAAA	AGGCCCCAG	GGGCCGACAG	CCAGCACCGC	5100
GAGGAAGCCC	ACCCACCCCA	CACACGACCA	CGGCAACCAA	ACCAGAACCC	AGACCACCCT	5160
GGGCCACCAG	CTCCCAGACT	CGGCCATCAC	CCCGCAGAAA	GGAAAGGCCA	CAACCCGCGC	5220
ACCCCAGCCC	CGATCCGGCG	GGGAGCCACC	CAACCCGAAC	CAGCACCCAA	GAGCGATCCC	5280
CGAAGGACCC	CCGAACCGCA	AAGGACATCA	GTATCCCACA	GCCTCTCCAA	GTCCCCCGGT	5340
CTCCTCCTCT	TCTCGAAGGG	ACCAAAAGAT	CAATCCACCA	CACCCGACGA	CACTCAACTC	5400
CCCACCCCTA	AAGGAGACAC	CGGGAATCCC	AGAATCAAGA	CTCATCCAAT	GTCCATCATG	5460
GGTCTCAAGG	TGAACGTCTC	TGCCATATTC	ATGGCAGTAC	TGTTAACTCT	CCAAACACCC	5520
ACCGGTCAAA	TCCATTGGGG	CAATCTCTCT	AAGATAGGGG	TGGTAGGAAT	AGGAAGTGCA	5580
AGCTACAAAG	TTATGACTCG	TTCCAGCCAT	CAATCATTAG	TCATAAAATT	AATGCCCAAT	5640
ATAACTCTCC	TCAATAACTG	CACGAGGGTA	GAGATTGCAG	AATACAGGAG	ACTACTGAGA	5700
ACAGTTTTGG	AACCAATTAG	AGATGCACTT	AATGCAATGA	CCCAGAATAT	AAGACCGGTT	5760
CAGAGTGTAG	CTTCAAGTAG	GAGACACAAG	AGATTTGCGG	GAGTAGTCCT	GGCAGGTGCG	5820
GCCCTAGGCG	TTGCCACAGC	TGCTCAGATA	ACAGCCGGCA	TTGCACTTCA	CCAGTCCATG	5880
CTGAACTCTC	AAGCCATCGA	CAATCTGAGA	GCGAGCCTGG	AAACTACTAA	TCAGGCAATT	5940
GAGGCAATCA	GACAAGCAGG	GCAGGAGATG	ATATTGGCTG	TTCAGGGTGT	CCAAGACTAC	6000
ATCAATAATG	AGCTGATACC	GTCTATGAAC	CAACTATCTT	GTGATTTAAT	CGGCCAGAAG	6060
CTCGGGCTCA	AATTGCTCAG	ATACTATACA	GAAATCCTGT	CATTATTTGG	CCCCAGCTTA	6120
CGGGACCCCA	TATCTGCGGA	GATATCTATC	CAGGCTTTGA	GCTATGCGCT	TGGAGGAGAC	6180
ATCAATAAGG	TGTTAGAAAA	GCTCGGATAC	AGTGGAGGTG	ATTTACTGGG	CATCTTAGAG	6240
AGCAGAGGAA	TAAAGGCCCG	GATAACTCAC	GTCGACACAG	AGTCCTACTT	CATTGTCCTC	6300
			AAGGGGGTGA			6360
GTCTCGTACA	ACATAGGCTC	TCAAGAGTGG	TATACCACTG	TGCCCAAGTA	TGTTGCAACC	6420

- 222 -

CAAGGGTACC	TTATCTCGAA	TTTTGATGAG	TCATCGTGTA	CTTTCATGCC	AGAGGGGACT	6480
GTGTGCAGCC	AAAATGCCTT	GTACCCGATG	AGTCCTCTGC	TCCAAGAATG	CCTCCGGGG	6540
TACACCAAGT	CCTGTGCTCG	TACACTCGTA	TCCGGGTCTT	TTGGGAACCG	GTTCATTTTA	6600
TCACAAGGGA	ACCTAATAGC	CAATTGTGCA	TCAATCCTTT	GCAAGTGTTA	CACAACAGGA	6660
ACGATCATTA	ATCAAGACCC	TGACAAGATC	CTAACATACA	TTGCTGCCGA	TAACTGCCCG	6720
GTAGTCGAGG	TGAACGGCGT	GACCATCCAA	GTCGGGAGCA	GGAGGTATCC	AGACGCTGTG	6780
TACTTGCACA	GAATTGACCT	CGGTCCTCCC	ATATTATTGG	AGAGGTTGGA	CGTAGGGACA	6840
AATCTGGGGA	ATGCAATTGC	TAAGTTGGAG	GATGCCAAGG	AATTGTTGGA	GTCATCGGAC	6900
CAGATATTGA	GGAGTATGAA	AGGTTTATCG	AGCACTTGCA	TAGTCTACAT	CCTGATTGCA	6960
GTGTGTCTTG	GAGGGTTGAT	AGGGATCCCC	GCTTTAATAT	GTTGCTGCAG	GGGGCGTTGT	7020
AACAAAAAGG	GAGAACAAGT	TGGTATGTCA	AGACCAGGCC	TAAAGCCTGA	TCTTACGGGA	7080
ACATCAAAAT	CCTATGTAAG	GTCGCTCTGA	TCCTCTACAA	CTCTTGAAAC	ACAAATGTCC	7140
CACAAGTCTC	CTCTTCGTCA	TCAAGCAACC	ACCGCACCCA	GCATCAAGCC	CACCTGAAAT	7200
TATCTCCGGC	TTCCCTCTGG	CCGAACAATA	TCGGTAGTTA	ATTAAAACTT	AGGGTGCAAG	7260
ATCATCCACA	ATGTCACCAC	AACGAGACCG	GATAAATGCC	TTCTACAAAG	ATAACCCCCA	7320
TCCCAAGGGA	AGTAGGATAG	TCATTAACAG	AGAACATCTT	ATGATTGATA	GACCTTATGT	7380
TTTGCTGGCT	GTTCTGTTTG	TCATGTTTCT	GAGCTTGATC	GGGTTGCTAG	CCATTGCAGG	7440
CATTAGACTT	CATCGGGCAG	CCATCTACAC	CGCAGAGATC	CATAAAAGCC	TCAGCACCAA	7500
TCTAGATGTA	ACTAACTCAA	TCGAGCATCA	GGTCAAGGAC	GTGCTGACAC	CACTCTTCAA	7560
AATCATCGGT	GATGAAGTGG	GCCTGAGGAC	ACCTCAGAGA	TTCACTGACC	TAGTGAAATT	7620
CATCTCTGAC	AAGATTAAAT	TCCTTAATCC	GGATAGGGAG	TACGACTTCA	GAGATCTCAC	7680
TTGGTGTATC	AACCCGCCAG	AGAGAATCAA	ATTGGATTAT	GATCAATACT	GTGCAGATGT	7740
GGCTGCTGAA	GAGCTCATGA	ATGCATTGGT	GAACTCAACT	CTACTGGAGA	CCAGAACAAC	7800
CAATCAGTTC	CTAGCTGTCT	CAAAGGGAAA	CTGCTCAGGG	CCCACTACAA	TCAGAGGTCA	7860
ATTCTCAAAC	ATGTCGCTGT	CCCTGTTAGA	CTTGTATTTA	GGTCGAGGTT	ACAATGTGTC	7920
ATCTATAGTC	ACTATGACAT	CCCAGGGAAT	GTATGGGGGA	ACTTACCTAG	TGGAAAAGCC	7980

- 223 -

TAATCTGAGC	agcaaaaggt	CAGAGTTGTC	ACAACTGAGC	ATGTACCGAG	TGTTTGAAGT	8040
AGGTGTTATC	AGAAATCCGG	GTTTGGGGGC	TCCGGTGTTC	CATATGACAA	ACTATCTTGA	8100
GCAACCAGTC	AGTAATGATC	TCAGCAACTG	TATGGTGGCT	TTGGGGGAGC	TCAAACTCGC	8160
AGCCCTTTGT	CACCGGGAAG	ATTCTATCAC	AATTCCCTAT	CAGGGATCAG	GGAAAGGTGT	8220
CAGCTTCCAG	CTCGTCAAGC	TAGGTGTCTG	GAAATCCCCA	ACCGACATGC	AATCCTGGGT	8280
CACCTTATCA	ACGGATGATC	CAGTGATAGA	CAGGCTTTAC	CTCTCATCTC	ACAGAGGTGT	8340
TATCGCTGAC	AATCAAGCAA	AATGGGCTGT	CCCGACAACA	CGAACAGATG	ACAAGTTGCG	8400
aatggagaca	TGCTTCCAAC	AGGCGTGTAA	GGGTAAAATC	CAAGCACTCT	GCGAGAATCC	8460
CGAGTGGGCA	CCATTGAAGG	ATAACAGGAT	TCCTTCATAC	GGGGTCTTGT	CTGTTGATCT	8520
GAGTCTGACA	GTTGAGCTTA	AAATCAAAAT	TGCTTCGGGA	TTCGGGCCAT	TGATCACACA	8580
CGGTTCAGGG	ATGGACCTAT	ACAAATCCAA	CCACAACAAT	GTGTATTGGC	TGACTATCCC	8640
ACCAATGAAG	AACCTAGCCT	TAGGTGTAAT	CAACACATTG	GAGTGGATAC	CGAGATTCAA	8700
GGTTAGTCCC	TACCTCTTCA	ATGTCCCAAT	TAAGGAAGCA	GGCGAAGACT	GCCATGCCCC	8760
AACATACCTA	CCTGCGGAGG	TGGATGGTGA	TGTCAAACTC	AGTTCCAATC	TGGTGATTCT	8820
ACCTGGTCAA	GATCTCCAAT	ATGTTTTGGC	AACCTACGAT	ACTTCCAGGG	TTGAACATGC	8880
TGTGGTTTAT	TACGTTTACA	GCCCAAGCCG	CTCATTTTCT	TACTTTTATC	CTTTTAGGTT	8940
GCCTATAAAG	GGGGTCCCCA	TCGAATTACA	AGTGGAATGC	TTCACATGGG	ACCAAAAACT	9000
CTGGTGCCGT	CACTTCTGTG	TGCTTGCGGA	CTCAGAATCT	GGTGGACATA	TCACTCACTC	9060
TGGGATGGTG	GGCATGGGAG	TCAGCTGCAC	AGTCACCCGG	GAAGATGGAA	CCAATCGCAG	9120
ATAGGGCTGC	TAGTGAACTA	ATCTCATGAT	GTCACCCAGA	CATCAGGCAT	ACCCACTAGT	9180
GTGAAATAGA	CATCAGAATT	AAGAAAAACG	TAGGGTCCAA	GTGGTTCCCC	GTTATGGACT	9240
CGCTATCTGT	CAACCAGATC	TTATACCCTG	AAGTTCACCT	AGATAGCCCG	ATAGTTACCA	9300
					CTGGAGGACC	
					CAAATGATTA	
					CCGGCCCACT	
CTCATATTCC	ATATCCAAAT	TGTAATCAGG	ATTTATTTAA	CATAGAAGAC	AAAGAGTCAA	9540

CGAGGAAGA1	CCGTGAACT	CTCAAAAAGG	GGAATTCGCT	GTACTCCAAI	GTCAGTGATA	9600
AGGTTTTCCA	ATGCTTAAGG	GACACTAACT	' CACGGCTTGG	CCTAGGCTCC	GAATTGAGGG	9660
AGGACATCAA	GGAGAAAGTT	ATTAACTTGG	GAGTTTACAT	GCACAGCTCC	CAGTGGTTTG	9720
AGCCCTTTCT	GTTTTGGTTT	ACAGTCAAGA	CTGAGATGAG	GTCAGTGATT	AAATCACAAA	9780
CCCATACTTG	CCATAGGAGG	AGACACACAC	CTGTATTCTT	CACTGGTAGT	TCAGTTGAGT	9840
TGCTAATCTC	TCGTGACCTT	GTTGCTATAA	TCAGTAAAGA	GTCTCAACAT	GTATATTACC	9900
TGACATTTGA	ACTGGTTTTG	ATGTATTGTG	ATGTCATAGA	GGGGAGGTTA	ATGACAGAGA	9960
CCGCTATGAC	TATTGATGCT	AGGTATACAG	AGCTTCTAGG	AAGAGTCAGA	TACATGTGGA	10020
AACTGATAGA	TGGTTTCTTC	CCTGCACTCG	GGAATCCAAC	TTATCAAATT	GTAGCCATGC	10080
TGGAGCCTCT	TTCACTTGCT	TACCTGCAGC	TGAGGGATAT	AACAGTAGAA	CTCAGAGGTG	10140
CTTTCCTTAA	CCACTGCTTT	ACTGAAATAC	ATGATGTTCT	TGACCAAAAC	GGGTTTTCTG	10200
ATGAAGGTAC	TTATCATGAG	TTAATTGAAG	CTCTAGATTA	CATTTTCATA	ACTGATGACA	10260
TACATCTGAC	AGGGGAGATT	TTCTCATTTT	TCAGAAGTTT	CGGCCACCCC	AGACTTGAAG	10320
CAGTAACGGC	TGCTGAAAAT	GTTAGGAAAT	ACATGAATCA	GCCTAAAGTC	ATTGTGTATG	10380
AGACTCTGAT	GAAAGGTCAT	GCCATATTTT	GTGGAATCAT	AATCAACGGC	TATCGTGACA	10440
GGCACGGAGG	CAGTTGGCCA	CCGCTGACCC	TCCCCCTGCA	TGCTGCAGAC	ACAATCCGGA	10500
ATGCTCAAGC	TTCAGGTGAA	GGGTTAACAC	ATGAGCAGTG	CGTTGATAAC	TGGAAATCTT	10560
TTGCTGGAGT	GAAATTTGGC	TGCTTTATGC	CTCTTAGCCT	GGATAGTGAT	CTGACAATGT	10620
ACCTAAAGGA	CAAGGCACTT	GCTGCTCTCC	AAAGGGAATG	GGATTCAGTT	TACCCGAAAG	10680
AGTTCCTGCG	TTACGACCCT	CCCAAGGGAA	CCGGGTCACG	GAGGCTTGTA	GATGTTTTCC	10740
TTAATGATTC	GAGCTTTGAC	CCATATGATG	TGATAATGTA	TGTTGTAAGT	GGAGCTTACC	10800
TCCATGACCC	TGAGTTCAAC	CTGTCTTACA	GCCTGAAAGA	AAAGGAGATC	AAGGAAACAG	10860
GTAGACTTTT	TGCTAAAATG	ACTTACAAAA	TGAGGGCATG	CCAAGTGATT	GCTGAAAATC	10920
TAATCTCAAA	CGGGATTGGC	AAATATTTTA	AGGACAATGG	GATGGCCAAG	GATGAGCACG	10980
ATTTGACTAA	GGCACTCCAC	ACTCTAGCTG	TCTCAGGAGT	CCCCAAAGAT	CTCAAAGAAA	11040
GTCACAGGGG	GGGGCCAGTC	TTAAAAACCT .	ACTCCCGAAG	CCCAGTCCAC	ACAAGTACCA	11100

- 225 -

GGAACGTGAG	AGCAGCAAAA	GGGTTTATAG	GGTTCCCTCA	AGTAATTCGG	CAGGACCAAG	11160
ACACTGATCA	TCCGGAGAAT	ATGGAAGCTT	ACGAGACAGT	CAGTGCATTT	ATCACGACTG	11220
ATCTCAAGAA	GTACTGCCTT	AATTGGAGAT	ATGAGACCAT	CAGCTTGTTT	GCACAGAGGC	11280
TAAATGAGAT	TTACGGATTG	CCCTCATTTT	TCCAGTGGCT	GCATAAGAGG	CTTGAGACCT	11340
CTGTCCTGTA	TGTAAGTGAC	CCTCATTGCC	CCCCGACCT	TGACGCCCAT	ATCCCGTTAT	11400
ATAAAGTCCC	CAATGATCAA	ATCTTCATTA	AGTACCCTAT	GGGAGGTATA	GAAGGGTATT	11460
GTCAGAAGCT	GTGGACCATC	AGCACCATTC	CCTATCTATA	CCTGGCTGCT	TATGAGAGCG	11520
GAGTAAGGAT	TGCTTCGTTA	GTGCAAGGGG	ACAATCAGAC	CATAGCCGTA	ACAAAAAGGG	11580
TACCCAGCAC	ATGGCCCTAC	AACCTTAAGA	AACGGGAAGC	TGCTAGAGTA	ACTAGAGATT	11640
ACTTTGTAAT	TCTTAGGCAA	AGGCTACATG	ATATTGGCCA	TCACCTCAAG	GCAAATGAGA	11700
CAATTGTTTC	ATCACATTTT	TTTGTCTATT	CAAAAGGAAT	ATATTATGAT	GGGCTACTTG	11760
TGTCCCAATC	ACTCAAGAGC	ATCGCAAGAT	GTGTATTCTG	GTCAGAGACT	ATAGTTGATG	11820
AAACAAGGGC	AGCATGCAGT	AATATTGCTA	CAACAATGGC	TAAAAGCATC	GAGAGAGGTT	11880
ATGACCGTTA	CCTTGCATAT	TCCCTGAACG	TCCTAAAAGT	GATACAGCAA	ATTCTGATCT	11940
CTCTTGGCTT	CACAATCAAT	TCAACCATGA	CCCGGGATGT	AGTCATACCC	CTCCTCACAA	12000
ACAACGACCT	CTTAATAAGG	ATGGCACTGT	TGCCCGCTCC	TATTGGGGG	ATGAATTATC	12060
TGAATATGAG	CAGGCTGTTT	GTCAGAAACA	TCGGTGATCC	AGTAACATCA	TCAATTGCTG	12120
ATCTCAAGAG	AATGATTCTC	GCCTCACTAA	TGCCTGAAGA	GACCCTCCAT	CAAGTAATGA	12180
CACAACAACC	GGGGGACTCT	TCATTCCTAG	ACTGGGCTAG	CGACCCTTAC	TCAGCAAATC	12240
TTGTATGTGT	CCAGAGCATC	ACTAGACTCC	TCAAGAACAT	AACTGCAAGG	TTTGTCCTGA	12300
TCCATAGTCC	AAACCCAATG	TTAAAAGGAT	TATTCCATGA	TGACAGTAAA	GAAGAGGACG	12360
AGGGACTGGC	GGCATTCCTC	ATGGACAGGC	ATATTATAGT	ACCTAGGGCA	GCTCATGAAA	12420
TCCTGGATCA	TAGTGTCACA	GGGGCAAGAG	AGTCTATTGC	AGGCATGCTG	GATACCACAA	12480
AAGGCCTGAT	TCGAGCCAGC	atgaggaagg	GGGGGTTAAC	CTCTCGAGTG	ATAACCAGAT	12540
TGTCCAATTA	TGACTATGAA	CAATTCAGAG	CAGGGATGGT	GCTATTGACA	GGAAGAAAGA	12600
GAAATGTCCT	CATTGACAAA	GAGTCATGTT	CAGTGCAGCT	GGCGAGAGCT	CTAAGAAGCC	12660

- 226 -

ATATGTGGGC GAGGCTAGCT CGAGGACGGC CTATTTACGG CCTTGAGGTC CCTGATGTAC	
	12720
TAGAATCTAT GCGAGGCCAC CTTATTCGGC GTCATGAGAC ATGTGTCATC TGCGAGTGTG	12780
GATCAGTCAA CTACGGATGG TTTTTTGTCC CCTCGGGTTG CCAACTGGAT GATATTGACA	12840
AGGAAACATC ATCCTTGAGA GTCCCATATA TTGGTTCTAC CACTGATGAG AGAACAGACA	12900
TGAAGCTTGC CTTCGTAAGA GCCCCAAGTC GATCCTTGCG ATCTGCTGTT AGAATAGCAA	12960
CAGTGTACTC ATGGGCTTAC GGTGATGATG ATAGCTCTTG GAACGAAGCC TGGTTGTTGG	13020
CTAGGCAAAG GGCCAATGTG AGCCTGGAGG AGCTAAGGGT GATCACTCCC ATCTCAACTT	13080
CGACTAATTT AGCGCATAGG TTGAGGGATC GTAGCACTCA AGTGAAATAC TCAGGTACAT	13140
CCCTTGTCCG AGTGGCGAGG TATACCACAA TCTCCAACGA CAATCTCTCA TTTGTCATAT	13200
CAGATAAGAA GGTTGATACT AACTTTATAT ACCAACAAGG AATGCTTCTA GGGTTGGGTG	13260
TTTTAGAAAC ATTGTTTCGA CTCGAGAAAG ATACCGGATC ATCTAACACG GTATTACATC	13320
TTCACGTCGA AACAGATTGT TGCGTGATCC CGATGATAGA TCATCCCAGG ATACCCAGCT	13380
CCCGCAAGCT AGAGCTGAGG GCAGAGCTAT GTACCAACCC ATTGATATAT GATAATGCAC	13440
CTTTAATTGA CAGAGATACA ACAAGGCTAT ACACCCAGAG CCATAGGAGG CACCTTGTGG	13500
ARTTTGTTAC ATGGTCCACA CCCCAACTAT ATCACATTTT AGCTAAGTCC ACAGCACTAT	13560
CTATGATTGA CCTGGTAACA AAATTTGAGA AGGACCATAT GAATGAAATT TCAGCTCTCA	13620
TAGGGGATGA CGATATCAAT AGTTTCATAA CTGAGTTTCT GCTCATAGAG CCAAGATTAT	13680
TCACTATCTA CTTGGGCCAG TGTGCGGCCA TCAATTGGGC ATTTGATGTA CATTATCATA	13740
GACCATCAGG GAAATATCAG ATGGGTGAGC TGTTGTCATC GTTCCTTTCT AGAATGAGCA	13800
AAGGAGTGTT TAAGGTGCTT GTCAATGCTC TAAGCCACCC AAAGATCTAC AAGAAATTCT	13860
GGCATTGTGG TATTATAGAG CCTATCCATG GTCCTTCACT TGATGCTCAA AACTTGCACA	13920
CAACTGTGTG CAACATGGTT TACACATGCT ATATGACCTA CCTCGACCTG TTGTTGAATG	13980
AAGAGTTAGA AGAGTTCACA TTTCTCTTGT GTGAAAGCGA CGAGGATGTA GTACCGGACA	14040
GATTCGACAA CATCCAGGCA AAACACTTAT GTGTTCTGGC AGATTTGTAC TGTCAACCAG	14100
GGGCCTGCCC ACCAATTCGA GGTCTAAGAC CGGTAGAGAA ATGTGCAGTT CTAACCGACC	14160
ATATCAAGGC AGAGGCTAGG TTATCTCCAG CAGGATCTTC GTGGAACATA AATCCAATTA	14220

- 227 -

TTGTAGACCA	TTACTCATGC	TCTCTGACTT	ATCTCCGGCG	AGGATCGATC	AAACAGATAA	14280
GATTGAGAGT	TGATCCAGGA	TTCATTTTCG	ACGCCCTCGC	TGAGGTAAAT	GTCAGTCAGC	14340
CAAAGATCGG	CAGCAACAAC	ATCTCAAATA	TGAGCATCAA	GGCTTTCAGA	CCCCCACACG	14400
ATGATGTTGC	AAAATTGCTC	AAAGATATCA	ACACAAGCAA	GCACAATCTT	CCCATTTCAG	14460
GGGGCAATCT	CGCCAATTAT	GAAATCCATG	CTTTCCGCAG	AATCGGGTTG	AACTCATCTG	14520
CTTGCTACAA	AGCTGTTGAG	ATATCAACAT	TAATTAGGAG	ATGCCTTGAG	CCAGGGGAGG	14580
ACGGCTTGTT	CTTGGGTGAG	GGATCGGGTT	CTATGTTGAT	CACTTATAAG	GAGATACTTA	14640
AACTAAACAA	GTGCTTCTAT	AATAGTGGGG	TTTCCGCCAA	TTCTAGATCT	GGTCAAAGGG	14700
AATTAGCACC	CTATCCCTCC	GAAGTTGGCC	TTGTCGAACA	CAGAATGGGA	GTAGGTAATA	14760
TTGTCAAAGT	GCTCTTTAAC	GGGAGGCCCG	AAGTCACGTG	GGTAGGCAGT	GTAGATTGCT	14820
TCAATTTCAT	AGTTAGTAAT	ATCCCTACCT	CTAGTGTGGG	GTTTATCCAT	TCAGATATAG	14880
AGACCTTGCC	TAACAAAGAT	ACTATAGAGA	AGCTAGAGGA	ATTGGCAGCC	ATCTTATCGA	14940
TGGCTCTGCT	CCTGGGCAAA	ATAGGATCAA	TACTGGTGAT	TAAGCTTATG	CCTTTCAGCG	15000
GGGATTTTGT	TCAGGGATTT	ATAAGTTATG	TAGGGTCTTA	TTATAGAGAA	GTGAACCTTG	15060
TATACCCTAG	ATACAGCAAC	TTCATATCTA	CTGAATCTTA	TTTGGTTATG	ACAGATCTCA	15120
AGGCTAACCG	GCTAATGAAT	CCTGAAAAGA	TTAAGCAGCA	GATAATTGAA	TCATCTGTGA	15180
GGACTTCACC	TGGACTTATA	GGTCACATCC	TATCCATTAA	GCAACTAAGC	TGCATACAAG	15240
CAATTGTGGG	AGACGCAGTT	AGTAGAGGTG	ATATCAATCC	TACTCTGAAA	AAACTTACAC	15300
CTATAGAGCA	GGTGCTGATC	AATTGCGGGT	TGGCAATTAA	CGGACCTAAG	CTGTGCAAAG	15360
AATTGATCCA	CCATGATGTT	GCCTCAGGGC	AAGATGGATT	GCTTAATTCT	ATACTCATCC	15420
TCTACAGGGA	GTTGGCAAGA	TTCAAAGACA	ACCGAAGAAG	TCAACAAGGG	ATGTTCCACG	15480
CTTACCCCGT	ATTGGTAAGT	AGCAGGCAAC	GAGAACTTAT	ATCTAGGATC	ACCCGCAAAT	15540
TTTGGGGGCA	CATTCTTCTT	TACTCCGGGA	ACAGAAAGTT	GATAAATAAG	TTTATCCAGA	15600
ATCTCAAGTC	CGGCTATCTG	ATACTAGACT	TACACCAGAA	TATCTTCGTT	AAGAATCTAT	15660
CCAAGTCAGA	GAAACAGATT	ATTATGACGG	GGGGTTTGAA	ACGTGAGTGG	GTTTTTAAGG	15720
TAACAGTCAA	GGAGACCAAA	GAATGGTATA	AGTTAGTCGG	ATACAGTGCC	CTGATTAAGG	15780

- 228 -

ACTAATTGAT TGAACTCCGG AACCCTAATC CTGCCCTAGG TGGTTAGGCA TTATTTGCAA 15840
TATATTAAAG AAAACTTTGA AAATACGAAG TTTCTATTCC CAGCTTTGTC TGGT 15894

- (2) INFORMATION FOR SEQ ID NO:16:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2183 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
  - Met Asp Ser Leu Ser Val Asn Gln Ile Leu Tyr Pro Glu Val His Leu 1 5 10 15
  - Asp Ser Pro Ile Val Thr Asn Lys Ile Val Ala Ile Leu Glu Tyr Ala 20 25 30
  - Arg Val Pro His Ala Tyr Ser Leu Glu Asp Pro Thr Leu Cys Gln Asn 35 40 45
  - Ile Lys His Arg Leu Lys Asn Gly Phe Ser Asn Gln Met Ile Ile Asn 50 55 60
  - Asn Val Glu Val Gly Asn Val Ile Lys Ser Lys Leu Arg Ser Tyr Pro 65 70 75 80
  - Ala His Ser His Ile Pro Tyr Pro Asn Cys Asn Gln Asp Leu Phe Asn 85 90 95
  - Ile Glu Asp Lys Glu Ser Thr Arg Lys Ile Arg Glu Leu Leu Lys Lys 100 105 110
  - Gly Asn Ser Leu Tyr Ser Lys Val Ser Asp Lys Val Phe Gln Cys Leu 115 120 125
  - Arg Asp Thr Asn Ser Arg Leu Gly Leu Gly Ser Glu Leu Arg Glu Asp 130 135 140
  - Ile Lys Glu Lys Val Ile Asn Leu Gly Val Tyr Met His Ser Ser Gln
    145 150 155 160
  - Trp Phe Glu Pro Phe Leu Phe Trp Phe Thr Val Lys Thr Glu Met Arg 165 170 175

- 229 -

Ser	Val	Ile	Lув 180	Ser	Gln	Thr	His	Thr 185	Сув	His	Arg	Arg	Arg 190	His	Thr
Pro	Val	Phe 195	Phe	Thr	Gly	Ser	Ser 200	Val	Glu	Leu	Leu	11e 205	Ser	Arg	Asp
Leu	Val 210	Ala	Ile	Ile	Ser	Lys 215	Glu	Ser	Gln	His	Val 220	Tyr	Tyr	Leu	Thr
Phe 225	Glu	Leu	Val	Leu	Met 230	Tyr	Сув	Asp	Val	Ile 235	Glu	Gly	Arg	Leu	Met 240
Thr	Glu	Thr	λla	Met 245	Thr	Ile	Авр	Ala	Arg 250	Tyr	Thr	Glu	Leu	Leu 255	Gly
Arg	Val	Arg	Tyr 260		Trp	Lys	Leu	11e 265	Asp	Gly	Phe	Phe	Pro 270	Ala	Leu
Gly	Asn	Pro 275	Thr	Tyr	Gln	Ile	Val 280	Ala	Met	Leu	Glu	Pro 285	Leu	Ser	Leu
Ala	Tyr 290		Gln	Leu	Arg	Asp 295	Ile	Thr	Val	Glu	Leu 300	Arg	Gly	Ala	Phe
Leu 305		His	Сув	Phe	Thr 310	Glu	Ile	His	Asp	Val 315	Leu	Авр	Gln	Asn	Gly 320
Phe	Ser	Авр	Glu	Gly 325		Tyr	His	Glu	Leu 330		Glu	Ala	Leu	Авр 335	Tyr
Ile	Phe	Ile	Thr 340		Asp	Ile	His	Leu 345	Thr	Gly	Glu	Ile	Phe 350	Ser	Phe
Phe	Arg	Ser 355		Gly	His	Pro	Arg 360		Glu	Ala	Val	Thr 365	Ala	Ala	Glu
Asn	Val 370		, Lys	Tyr	Met	375	Gln	Pro	Lys	Val	380	Val	Туг	Glu	Thr
Leu 385		: Lys	Gly	, His	390		Phe	Сув	Gly	7 Ile 395		Ile	Asn	Gly	Tyr 400
Arg	, Asp	Arg	g His	6 Gly		Ser	Trp	Pro	Pro 410		Thr	Lev	Pro	415	His
Ala	Ala	a Ası	7 Thi		Arg	Asr	n Ala	425		s Ser	Gly	/ Glu	1 Gly 430	Leu )	Thr
His	Glu	ı Glı 43		s Val	L Ası	) Ası	1 Trp	Lys	s Sei	Phe	Ala	Gl <sub>3</sub> 445	y Val	Lya	Phe
Gly	r Cys	s Pho	e Me	t Pro	Let	se:	r Let	Asp	Sei	r Asy	Let	ı Thi	r Met	туз	r Leu

- 230 -

	45	0				45	5				46	0			
Lys 465	As <sub>l</sub>	p Ly	s Al	a Lev	470	Ala	a Le	u Gl	n Ar	g Gl:	u Tr <sub>i</sub>	P Ası	Se:	r Vai	1 Tyr 480
Pro	Ly	3 Gl	u Ph	e Leu 485	Arg	Туз	. yal	Pr	0 Pro	o Lya	Gl <sub>y</sub>	/ Thi	Gly	7 Sex	r Arg
Arg	Let	ı Val	L Asj 500	y Val	Phe	Lev	ı Asr	1 Авј 50!	9 Sei 5	r Ser	Phe	Asp	Pro 510		Asp
Val	Ile	Met 515	Туз	Val	Val	Ser	Gly 520	Ala	ту:	r Leu	His	Asp 525		Glu	Phe
Asn	Leu 530	Ser	Туг	Ser	Leu	Lys 535	Glu	Lys	Glu	ı Ile	Lys 540		Thr	Gly	Arg
Leu 5 <b>45</b>	Phe	Ala	Lys	Met	Thr 550	Tyr	Lys	Met	Arg	7 Ala 555	Сув	Gln	Val	Ile	Ala 560
				Ser 565					570					575	
			300					585					590		
Val	Ser	Gly 595	Val	Pro	Lys	Asp	Leu 600	Lys	Glu	Ser	His	Arg 605	Gly	Gly	Pro
	010			Tyr		615					620				
<b>V</b>				Lys	630					635					640
				Авр 645					650					655	
			990	Thr				665					670		
		4,3		Ser :			680					685			
Leu i	Pro 690	Ser	Phe	Phe (	Gln :	rp 595	Leu :	His	Lys	Arg	Leu 700	Glu '	Thr	Ser	Val
Leu 7 705				•	10					715					720
Pro I	en ,	Tyr	Lys	Val I 725	Pro A	lsn i	Asp (	Gln	Ile 730	Phe :	Ile	Lys :		Pro 1	Met

PCT/US97/16718 WO 98/13501

- 231 -

Gly	Gly	Ile	Glu 740	Gly	Tyr	Сув	Gln	Lys 745	Leu	Trp	Thr	Ile	Ser 750	Thr	Ile
Pro	Tyr	Leu 755	Týx	Leu	Ala	Ala	Tyr 760	Glu	Ser	Gly	Val	<b>Arg</b> 765	Ile	Ala	Ser
Leu	Val 770	Gln	Gly	Asp	λsn	Gln 775	Thr	Ile	Ala	Val	Thr 780	Lys	Arg	Val	Pro
Ser 785	Thr	Trp	Pro	Tyr	Asn 790	Leu	Lys	Lys	Arg	Glu 795	Ala	Ala	Arg	Val	Thr 800
			Phe	805					810					815	
			Ala 820					825					830		
		835	Ile				840					845			
	850		Arg			855					860				
865			Сув		870					875					880
			Asp	885					890					895	
			11e 900					905					910		
		915					920					925			Ile
	930					935					940				Asn
945					950					955					Ser 960
				965					970					975	Glu
			980					985					990		Leu
Asp	Trp	Ala 995		Asp	Pro	Tyr	Ser 100		. Asn	Leu	Val	Сув 100		Gln	Ser

- 232 -

- Ile Thr Arg Leu Leu Lys Asn Ile Thr Ala Arg Phe Val Leu Ile His 1010 1015 1020
- Ser Pro Asn Pro Met Leu Lys Gly Leu Phe His Asp Asp Ser Lys Glu 1025 1030 1035 1040
- Glu Asp Glu Gly Leu Ala Ala Phe Leu Met Asp Arg His Ile Ile Val 1045 1050 1055
- Pro Arg Ala Ala His Glu Ile Leu Asp His Ser Val Thr Gly Ala Arg
  1060 1065 1070
- Glu Ser Ile Ala Gly Met Leu Asp Thr Thr Lys Gly Leu Ile Arg Ala 1075 1080 1085
- Ser Met Arg Lys Gly Gly Leu Thr Ser Arg Val Ile Thr Arg Leu Ser 1090 1095 1100
- Asn Tyr Asp Tyr Glu Gln Phe Arg Ala Gly Met Val Leu Leu Thr Gly 1105 1110 1115 1120
- Arg Lys Arg Asn Val Leu Ile Asp Lys Glu Ser Cys Ser Val Gln Leu 1125 1130 1135
- Ala Arg Ala Leu Arg Ser His Met Trp Ala Arg Leu Ala Arg Gly Arg 1140 1145 1150
- Pro Ile Tyr Gly Leu Glu Val Pro Asp Val Leu Glu Ser Met Arg Gly 1155 1160 1165
- His Leu Ile Arg Arg His Glu Thr Cys Val Ile Cys Glu Cys Gly Ser 1170 1175 1180
- Val Asn Tyr Gly Trp Phe Phe Val Pro Ser Gly Cys Gln Leu Asp Asp 1185 1190 1195 1200
- Ile Asp Lys Glu Thr Ser Ser Leu Arg Val Pro Tyr Ile Gly Ser Thr 1205 1210 1215
- Thr Asp Glu Arg Thr Asp Met Lys Leu Ala Phe Val Arg Ala Pro Ser 1220 1225 1230
- Arg Ser Leu Arg Ser Ala Val Arg Ile Ala Thr Val Tyr Ser Trp Ala 1235 1240 1245
- Tyr Gly Asp Asp Ser Ser Trp Asn Glu Ala Trp Leu Leu Ala Arg 1250 1260
- Gln Arg Ala Asn Val Ser Leu Glu Glu Leu Arg Val Ile Thr Pro Ile 1265 1270 1275 1280
- Ser Thr Ser Thr Asn Leu Ala His Arg Leu Arg Asp Arg Ser Thr Gln

- 233 -

				128	5				129	0				129	5
Val	Lув	Tyr	Ser 130	Gly 0	Thr	Ser	Leu	Val 130		Val	Ala	Arg	Tyr 131		Thr
Ile	Ser	Asn 131		Asn	Leu	Ser	Phe 132		Ile	Ser	Авр	Lys 132	-	Val	Авр
Thr	Asn 133		Ile	Tyr	Gln	Gln 133	_	Met	Leu	Leu	Gly 1340		Gly	Val	Leu
Glu 134		Leu	Phe	Arg	Leu 135(		Lys	Двр	, Thr	Gly 135		Ser	Asn	Thr	Val 1360
Leu	His	Leu	His	Val 1369		Thr	Asp	Сув	Сув 137		Ile	Pro	Met	11e 1379	_
His	Pro	Arg	Ile 1380	Pro	Ser	Ser	Arg	Lys 138		Glu	Leu	Arg	Ala 1390		Leu
Сув	Thr	Авп 139		Leu	Ile	Tyr	Авр 1400		Ala	Pro	Leu	Ile 1405		Arg	Asp
Thr	Thr 1410		Leu	Tyr	Thr	Gln 141		His	Arg	Arg	His 1420		Val	Glu	Phe
Val 1425		Trp	Ser	Thr	Pro 1430		Leu	Tyr	His	Ile 1435		Ala	Lys	Ser	Thr 1440
Ala	Leu	Ser	Met	Ile 1445		Leu	Val	Thr	Lys 1450		Glu	Lys	Авр	His 1455	
Asn	Glu	Ile	Ser 1460	Ala	Leu	Ile	Gly	Авр 1465		Asp	Ile	Asn	Ser 1470		Ile
Thr	Glu	Phe 1475		Leu	Ile	Glu	Pro 1480		Leu	Phe	Thr	Ile 1485	_	Leu	Gly
Gln	Сув 1490		Ala	Ile	Asn	Trp 1495		Phe	Asp	Val	Н1s 1500	_	His	Arg	Pro
Ser 1505		Lys	Tyr	Gln	Met 1510		Glu	Leu	Leu	Ser 1515		Phe	Leu	Ser	Arg 1520
Met	Ser	Lys	Gly	Val 1525		Lys	Val	Leu	Val 1530		Ala	Leu	Ser	His 1535	
Lys	Ile	Tyr	Lys 1540	Lys	Phe	Trp		Сув 1545	_	Ile	Ile		Pro 1550		His
Gly		Ser 1555		Asp	Ala		Asn 1560		His	Thr		Val 1565	_	Asn :	Met

- Val Tyr Thr Cys Tyr Met Thr Tyr Leu Asp Leu Leu Leu Asn Glu Glu 1570 1580
- Leu Glu Glu Phe Thr Phe Leu Leu Cys Glu Ser Asp Glu Asp Val Val 1585 1590 1595 1600
- Pro Asp Arg Phe Asp Asn Ile Gln Ala Lys His Leu Cys Val Leu Ala 1605 1610 1615
- Asp Leu Tyr Cys Gln Pro Gly Ala Cys Pro Pro Ile Arg Gly Leu Arg 1620 1625 1630
- Pro Val Glu Lys Cys Ala Val Leu Thr Asp His Ile Lys Ala Glu Ala 1635 1640 1645
- Arg Leu Ser Pro Ala Gly Ser Ser Trp Asn Ile Asn Pro Ile Ile Val 1650 1655 1660
- Asp His Tyr Ser Cys Ser Leu Thr Tyr Leu Arg Arg Gly Ser Ile Lys 1665 1670 1675 1680
- Gin Ile Arg Leu Arg Val Asp Pro Gly Phe Ile Phe Asp Ala Leu Ala 1685 1690 1695
- Glu Val Asn Val Ser Gln Pro Lys Ile Gly Ser Asn Asn Ile Ser Asn 1700 1705 1710
- Met Ser Ile Lys Ala Phe Arg Pro Pro His Asp Asp Val Ala Lys Leu 1715 1720 1725
- Leu Lys Asp Ile Asn Thr Ser Lys His Asn Leu Pro Ile Ser Gly Gly 1730 1740
- Asn Leu Ala Asn Tyr Glu Ile His Ala Phe Arg Arg Ile Gly Leu Asn 1745 1750 1755 1760
- Ser Ser Ala Cys Tyr Lys Ala Val Glu Ile Ser Thr Leu Ile Arg Arg 1765 1770 1775
- Cys Leu Glu Pro Gly Glu Asp Gly Leu Phe Leu Gly Glu Gly Ser Gly 1780 1785 1790
- Ser Met Leu Ile Thr Tyr Lys Glu Ile Leu Lys Leu Asn Lys Cys Phe 1795 1800 1805
- Tyr Asn Ser Gly Val Ser Ala Asn Ser Arg Ser Gly Gln Arg Glu Leu 1810 1815 1820
- Ala Pro Tyr Pro Ser Glu Val Gly Leu Val Glu His Arg Met Gly Val 1825 1830 1835 1840

- Gly Asn Ile Val Lys Val Leu Phe Asn Gly Arg Pro Glu Val Thr Trp 1845 1850 1855
- Val Gly Ser Val Asp Cys Phe Asn Phe Ile Val Ser Asn Ile Pro Thr 1860 1865 1870
- Ser Ser Val Gly Phe Ile His Ser Asp Ile Glu Thr Leu Pro Asn Lys 1875 1880 1885
- Asp Thr Ile Glu Lys Leu Glu Glu Leu Ala Ala Ile Leu Ser Met Ala 1890 1895 1900
- Leu Leu Cly Lys Ile Gly Ser Ile Leu Val Ile Lys Leu Met Pro 1905 1910 1915 1920
- Phe Ser Gly Asp Phe Val Gln Gly Phe Ile Ser Tyr Val Gly Ser Tyr 1925 1930 1935
- Tyr Arg Glu Val Asn Leu Val Tyr Pro Arg Tyr Ser Asn Phe Ile Ser 1940 1945 1950
- Thr Glu Ser Tyr Leu Val Met Thr Asp Leu Lys Ala Asn Arg Leu Met 1955 1960 1965
- Asn Pro Glu Lys Ile Lys Gln Gln Ile Ile Glu Ser Ser Val Arg Thr 1970 1975 1980
- Ser Pro Gly Leu Ile Gly His Ile Leu Ser Ile Lys Gln Leu Ser Cys 1985 1990 1995 2000
- Ile Gln Ala Ile Val Gly Asp Ala Val Ser Arg Gly Asp Ile Asn Pro 2005 2010 2015
- Thr Leu Lys Lys Leu Thr Pro Ile Glu Gln Val Leu Ile Asn Cys Gly 2020 2025 2030
- Leu Ala Ile Asn Gly Pro Lys Leu Cys Lys Glu Leu Ile His His Asp 2035 2040 2045
- Val Ala Ser Gly Gln Asp Gly Leu Leu Asn Ser Ile Leu Ile Leu Tyr 2050 2055 2060
- Arg Glu Leu Ala Arg Phe Lys Asp Asn Arg Ser Gln Gln Gly Met 2065 2070 2075 2080
- Phe His Ala Tyr Pro Val Leu Val Ser Ser Arg Gln Arg Glu Leu Ile 2085 2090 2095
- Ser Arg Ile Thr Arg Lys Phe Trp Gly His Ile Leu Leu Tyr Ser Gly 2100 2105 2110
- Asn Arg Lys Leu Ile Asn Lys Phe Ile Gln Asn Leu Lys Ser Gly Tyr

- 236 -

2115 2120 2125

Leu Ile Leu Asp Leu His Gln Asn Ile Phe Val Lys Asn Leu Ser Lys 2130 2135 2140

Ser Glu Lys Gln Ile Ile Met Thr Gly Gly Leu Lys Arg Glu Trp Val 2145 2150 2155 2160

Phe Lys Val Thr Val Lys Glu Thr Lys Glu Trp Tyr Lys Leu Val Gly
2165 2170 2175

Tyr Ser Ala Leu Ile Lys Asp 2180

- (2) INFORMATION FOR SEQ ID NO:17:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 15462 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: RNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ACCAAACAAG AGAAGAAACT TGTCTGGGAA TATAAATTTA ACTTTAAATT AACTTAGGAT 60 TARAGACATT GACTAGAAGG TCAAGAAAAG GGAACTCTAT AATTTCAAAA ATGTTGAGCC 120 TATTTGATAC ATTTAATGCA CGTAGGCAAG AAAACATAAC AAAATCAGCC GGTGGAGCTA 180 TCATTCCTGG ACAGAAAAAT ACTGTCTCTA TATTCGCCCT TGGACCGACA ATAACTGATG 240 ATANTGAGAN ANTGACATTA GCTCTTCTAT TTCTATCTCA TTCACTAGAT ANTGAGANAC 300 AACATGCACA AAGGGCAGGG TTCTTGGTGT CTTTATTGTC AATGGCTTAT GCCAATCCAG 360 AGCTCTACCT AACAACAAAT GGAAGTAATG CAGATGTCAA GTATGTCATA TACATGATTG 420 AGAAAGATCT AAAACGGCAA AAGTATGGAG GATTTGTGGT TAAGACGAGA GAGATGATAT 480 ATGAAAAGAC AACTGATTGG ATATTTGGAA GTGACCTGGA TTATGATCAG GAAACTATGT 540 TGCAGAACGG CAGGAACAAT TCAACAATTG AAGACCTTGT CCACACATTT GGGTATCCAT 600 CATGTTTAGG AGCTCTTATA ATACAGATCT GGATAGTTCT GGTCAAAGCT ATCACTAGTA 660 TCTCAGGGTT AAGAAAAGGC TTTTCACCC GATTGGAAGC TTTCAGACAA GATGGAACAG 720

- 237 -

TGCAGGCAGG	GCTGGTATTG	AGCGGTGACA	CAGTGGATCA	GATTGGGTCA	ATCATGCGGT	780
CTCAACAGAG	CTTGGTAACT	CTTATGGTTG	AAACATTAAT	AACAATGAAT	ACCAGCAGAA	840
ATGACCTCAC	AACCATAGAA	AAGAATATAC	AAATTGTTGG	CAACTACATA	AGAGATGCAG	900
GTCTCGCTTC	ATTCTTCAAT	ACAATCAGAT	ATGGAATTGA	GACCAGAATG	GCAGCTTTGA	960
CTCTATCCAC	TCTCAGACCA	GATATCAATA	GATTAAAAGC	TTTGATGGAA	CTGTATTTAT	1020
CAAAGGGACC	ACGCGCTCCT	TTCATCTGTA	TCCTCAGAGA	TCCTATACAT	GGTGAGTTCG	1080
CACCAGGCAA	CTATCCTGCC	ATATGGAGCT	ATGCAATGGG	GGTGGCAGTT	GTACAAAATA	1140
GAGCCATGCA	ACAGTATGTG	ACGGGAAGAT	CATATCTAGA	CATTGATATG	TTCCAGCTAG	1200
GACAAGCAGT	AGCACGTGAT	GCCGAAGCTC	AAATGAGCTC	AACACTGGAA	GATGAACTTG	1260
GAGTGACACA	CGAATCTAAA	GAAAGCTTGA	AGAGACATAT	AAGGAACATA	AACAGTTCAG	1320
AGACATCTTT	CCACAAACCG	ACAGGTGGAT	CAGCCATAGA	GATGGCAATA	GATGAAGAGC	1380
CAGAACAATT	CGAACATAGA	GCAGATCAAG	AACAAAATGG	AGAACCTCAA	TCATCCATAA	1440
TTCAATATGC	CTGGGCAGAA	GGAAATAGAA	GCGATGATCA	GACTGAGCAA	GCTACAGAAT	1500
CTGACAATAT	CAAGACCGAA	CAACAAAACA	TCAGAGACAG	ACTAAACAAG	AGACTCAACG	1560
ACAAGAAGAA	ACAAAGCAGT	CAACCACCCA	CTAATCCCAC	AAACAGAACA	AACCAGGACG	1620
aaatagatga	TCTGTTTAAC	GCATTTGGAA	GCAACTAATC	GAATCAACAT	TTTAATCTAA	1680
ATCAATAATA	AATAAGAAAA	ACTTAGGATT	AAAGAATCCT	ATCATACCGG	AATATAGGGT	1740
GGTAAATTTA	GAGTCTGCTT	GAAACTCAAT	CAATAGAGAG	TTGATGGAAA	GCGATGCTAA	1800
AAACTATCAA	ATCATGGATT	CTTGGGAAGA	GGAATCAAGA	GATAAATCAA	CTAATATCTC	1860
CTCGGCCCTC	AACATCATTG	AATTCATACT	CAGCACCGAC	CCCCAAGAAG	ACTTATCGGA	1920
AAACGACACA	ATCAACACAA	GAACCCAGCA	ACTCAGTGCC	ACCATCTGTC	AACCAGAAAT	1980
CAAACCAACA	GAAACAAGTG	AGAAAGATAG	TGGATCAACT	GACAAAAATA	GACAGTCCGG	2040
GTCATCACAC	GAATGTACAA	CAGAAGCAAA	AGATAGAAAT	ATTGATCAGG	AAACTGTACA	2100
GAGAGGACCT	GGGAGAAGAA	GCAGCTCAGA	TAGTAGAGCT	GAGACTGTGG	TCTCTGGAGG	2160
AATCCCCAGA	AGCATCACAG	ATTCTAAAAA	TGGAACCCAA	AACACGGAGG	ATATTGATCT	2220
CAATGAAATT	DETABAAABA	ATAAGGACTC	TATTGAGGGG	AAAATGCGAC	AATCTGCAAA	2280

TGTTCCAAGC	GAGATATCAG	GAAGTGATGA	CATATTTACA	ACAGAACAAA	GTAGAAACAG	2340
TGATCATGGA	AGAAGCCTGG	AATCTATCAG	TACACCTGAT	ACAAGATCAA	TAAGTGTTGT	2400
TACTGCTGCA	ACACCAGATG	atgaagaaga	AATACTAATG	aaaaatagta	GGACAAAGAA	2460
AAGTTCTTCA	ACACATCAAG	AAGATGACAA	AAGAATTAAA	AAAGGGGGAA	AAGGGAAAGA	2520
CTGGTTTAAG	AAATCAAAAG	ATACCGACAA	CCAGATACCA	ACATCAGACT	ACAGATCCAC	2580
ATCAAAAGGG	CAGAAGAAAA	TCTCAAAGAC	AACAACCACC	AACACCGACA	CAAAGGGGCA	2640
AACAGAAATA	CAGACAGAAT	CATCAGAAAC	ACAATCCTCA	TCATGGAATC	TCATCATCGA	2700
CAACAACACC	GACCGGAACG	AACAGACAAG	CACAACTCCT	CCAACAACAA	CTTCCAGATC	2760
AACTTATACA	AAAGAATCGA	TCCGAACAAA	CTCTGAATCC	AAACCCAAGA	CACAAAAGAC	2820
AAATGGAAAG	GAAAGGAAGG	ATACAGAAGA	GAGCAATCGA	TTTACAGAGA	GGGCAATTAC	2880
TCTATTGCAG	AATCTTGGTG	TAATTCAATC	CACATCAAAA	CTAGATTTAT	ATCAAGACAA	2940
ACGAGTTGTA	TGTGTAGCAA	ATGTACTAAA	CAATGTAGAT	ACTGCATCAA	AGATAGATTT	3000
CCTGGCAGGA	TTAGTCATAG	GGGTTTCAAT	GGACAACGAC	ACAAAATTAA	CACAGATACA	3060
AAATGAAATG	CTAAACCTCA	AAGCAGATCT	AAAGAAAATG	GACGAATCAC	ATAGAAGATT	3120
GATAGAAAAT	CAAAGAGAAC	AACTGTCATT	GATCACGTCA	CTAATTTCAA	ATCTCAAAAT	3180
TATGACTGAG	AGAGGAGGAA	AGAAAGACCA	AAATGAATCC	aatgagagag	TATCCATGAT	3240
CAAAACAAAA	TTGAAAGAAG	AAAAGATCAA	GAAGACCAGG	TTTGACCCAC	TTATGGAGGC	3300
ACAAGGCATT	GACAAGAATA	TACCCGATCT	ATATCGACAT	GCAGGAGATA	CACTAGAGAA	3360
CGATGTACAA	GTTAAATCAG	AGATATTAAG	TTCATACAAT	GAGTCAAATG	CAACAAGACT	3420
AATACCCAAA	AAAGTGAGCA	GTACAATGAG	ATCACTAGTT	GCAGTCATCA	ACAACAGCAA	3480
TCTCTCACAA	AGCACAAAAC	AATCATACAT	AAACGAACTC	AAACGTTGCA	AAAATGATGA	3540
AGAAGTATCT	GAATTAATGG	ACATGTTCAA	TGAAGATGTC	AACAATTGCC	AATGATCCAA	3600
CAAAGAAACG	ACACCGAACA	AACAGACAAG	AAACAACAGT	AGATCAAAAC	CTGTCAACAC	3660
ACACAAAATC	AAGCAGAATG	AAACAACAGA	TATCAATCAA	TATACAAATA	AGAAAAACTT	3720
AGGATTAAAG	AATAAATTAA	TCCTTGTCCA	AAATGAGTAT	AACTAACTCT	GCAATATACA	3780
CATTCCCAGA	ATCATCATTC	TCTGAAAATG	GTCATATAGA	ACCATTACCA	CTCAAAGTCA	3840

- 239 -

ATGAACAGAG	GAAAGCAGT	CCCCACATT	GAGTTGCCAA	GATCGGAAAT	CCACCAAAAC .	3900
ACGGATCCC	GTATTTAGAT	GTCTTCTTAC	: TCGGCTTCTT	CGAGATGGAA	CGAATCAAAG	3960
ACAAATACGG	GAGTGTGAAT	GATCTCGACA	GTGACCCGAG	TTACAAAGTT	TGTGGCTCTG	4020
GATCATTACO	AATCGGATTG	GCTAAGTACA	CTGGGAATGA	CCAGGAATTG	TTACAAGCCG	4080
CAACCAAACT	GGATATAGAA	GTGAGAAGAA	CAGTCAAAGC	GAAAGAGATG	GTTGTTTACA	4140
CGGTACAAA	TATAAAACCA	GAACTGTACO	CATGGTCCAA	TAGACTAAGA	AAAGGAATGC	4200
TGTTCGATGC	CAACAAAGTT	GCTCTTGCTC	CTCAATGTCT	TCCACTAGAT	AGGAGCATAA	4260
AATTTAGAGT	AATCTTCGTG	AATTGTACGG	CAATTGGATC	AATAACCTTG	TTCAAAATTC	4320
CTAAGTCAAT	GGCATCACTA	TCTCTACCCA	ACACAATATC	AATCAATCTG	CAGGTACACA	4380
TAAAAACAGG	GGTTCAGACT	GATTCTAAAG	GGATAGTTCA	AATTTTGGAT	GAGAAAGGCG	4440
AAAAATCACT	GAATTTCATG	GTCCATCTCG	GATTGATCAA	AAGAAAAGTA	GGCAGAATGT	4500
ACTCTGTTGA	ATACTGTAAA	CAGAAAATCG	AGAAAATGAG	ATTGATATTT	TCTTTAGGAC	4560
TAGTTGGAGG	AATCAGTCTT	CATGTCAATG	CAACTGGGTC	CATATCAAAA	ACACTAGCAA	4620
			ATCCTTTAAT			4680
			TTACAAGAGT			4740
			ATATTATTGC			4800
			CCGGACGTAT			4860
			GTCAATACCA			4920
			AGTCAAATAG			4980
			TCCAACTCAC			5040
			AATGCCAACT			5100
			AGATATCACA			5160
•					GATATCTAAT	5220
					AGATCAAGCA	5280
			TGAAAACACT		GATTACAGAA	5340
	GINACCAMIC	AAGAATUUAA	IGAAAACACT	GATCCCAGAA	CAAAACGATT	5400

- 240 -

CTTTGGAGG	GTAATTGGA	CCATTGCTCT	GGGAGTAGCA	ACCTCAGCAC	AAATTACAGC	5460
GGCAGTTGCT	CTGGTTGAAG	CCAAGCAGGG	AAGATCAGAC	ATCGAAAAA	CTCAAAGAAGC	5520
AATTAGGGAC	: ACAAACAAAG	CAGTGCAGTC	: AGTTCAGAGO	: TCCATAGGAA	ATTTAATAGT	5580
AGCAATTAAA	TCAGTCCAGG	ATTATGTTA	CAAAGAAATO	GTGCCATCGA	TTGCGAGGCT	5640
AGGTTGTGAA	GCAGCAGGAC	TTCAATTAGG	AATTGCATTA	ACACAGCATT	ACTCAGAATT	5700
AACAAACATA	TTTGGTGATA	ACATAGGATO	GTTACAAGAA	AAAGGAATAA	AATTACAAGG	5760
TATAGCATCA	TTATACCGCA	CAAATATCAC	AGAAATATTC	ACAACATCAA	CAGTTGATAA	5820
ATATGATATO	TATGATCTGT	TATTTACAGA	ATCAATAAAG	GTGAGAGTTA	TAGATGTTGA	5880
CTTGAATGAT	TACTCAATCA	CCCTCCAAGT	CAGACTCCCT	TTATTAACTA	GGCTGCTGAA	5940
CACTCAGATC	TACAAAGTAG	ATTCCATATC	ATATAACATC	CAAAACAGAG	AATGGTATAT	6000
CCCTCTTCCC	AGCCATATCA	TGACGAAAGG	GGCATTTCTA	GGTGGAGCAG	ACGTCAAAGA	6060
ATGTATAGAA	GCATTCAGCA	GCTATATATG	CCCTTCTGAT	CCAGGATTTG	TATTAAACCA	6120
TGAAATAGAG	AGCTGCTTAT	CAGGAAACAT	ATCCCAATGT	CCAAGAACAA	CGGTCACATC	6180
AGACATTGTT	CCAAGATATG	CATTTGTCAA	TGGAGGAGTG	GTTGCAAACT	GTATAACAAC	6240
CACCTGTACA	TGCAACGGAA	TTGGTAATAG	AATCAATCAA	CCACCTGATC	AAGGAGTAAA	6300
AATTATAACA	CATAAAGAAT	GTAGTACAAT	AGGTATCAAC	GGAATGCTGT	TCAATACAAA	6360
TAAAGAAGGA	ACTCTTGCAT	TCTATACACC	AAATGATATA	ACACTAAACA	ATTCTGTTGC	6420
ACTTGATCCA	ATTGACATAT	CAATCGAGCT	CAACAAGGCC	AAATCAGATC	TAGAAGAATC	6480
AAAAGAATGG	ATAAGAAGGT	CAAATCAAAA	ACTAGATTCT	ATTGGAARTT	GGCATCAATC	6540
TAGCACTACA	ATCATAATTA	TTTTGATAAT	GATCATTATA	TTGTTTATAA	TTAATATAAC	6600
GATAATTACA	ATTGCAATTA	AGTATTACAG	AATTCAAAAG	AGAAATCGAG	TGGATCAAAA	6660
TGACAAGCCA	TATGTACTAA	CAAACAAATA	ACATATCTAC	AGATCATTAG	ATATTAAAAT	6720
TATAAAAAAC	TTAGGAGTAA	AGTTACGCAA	TCCAACTCTA	CTCATATAAT	TGAGGAAGGA	6780
CCCAATAGAC	AAATCCAAAT	TCGAGATGGA	ATACTGGAAG	CATACCAATC	ACGGAAAGGA	6840
TGCTGGTAAT	GAGCTGGAGA	CGTCTATGGC	TACTCATGGC	AACAAGCTCA	CTAATAAGAT	6900
AATATACATA	TTATGGACAA	TAATCCTGGT	GTTATTATCA	ATAGTCTTCA	TCATAGTGCT	6960

- 241 -

AATTAATTCC	ATCAAAAGTG	AAAAGGCCCA	CGAATCATTG	CTGCAAGACA	TAAATAATGA	7020
GTTTATGGAA	ATTACAGAAA	AGATCCAAAT	GGCATCGGAT	AATACCAATG	ATCTAATACA	7080
GTCAGGAGTG	AATACAAGGC	TTCTTACAAT	TCAGAGTCAT	GTCCAGAATT	ACATACCAAT	7140
ATCATTGACA	CAACAGATGT	CAGATCTTAG	GAAATTCATT	AGTGAAATTA	CAATTAGAAA	7200
TGATAATCAA	GAAGTGCTGC	CACAAAGAAT	AACACATGAT	GTAGGTATAA	AACCTTTAAA	7260
TCCAGATGAT	TTTTGGAGAT	GCACGTCTGG	TCTTCCATCT	TTAATGAAAA	CTCCAAAAAT	7320
AAGGTTAATG	CCAGGGCCGG	GATTATTAGC	TATGCCAACG	ACTGTTGATG	GCTGTGTTAG	7380
AACTCCGTCT	TTAGTTATAA	ATGATCTGAT	TTATGCTTAT	ACCTCAAATC	TAATTACTCG	7440
AGGTTGTCAG	GATATAGGAA	AATCATATCA	AGTCTTACAG	ATAGGGATAA	TAACTGTAAA	7500
CTCAGACTTG	GTACCTGACT	TAAATCCTAG	GATCTCTCAT	ACCTTTAACA	TAAATGACAA	7560
TAGGAAGTCA	TGTTCTCTAG	CACTCCTAAA	TACAGATGTA	TATCAACTGT	GTTCAACTCC	7620
CAAAGTTGAT	GAAAGATCAG	ATTATGCATC	ATCAGGCATA	GAAGATATTG	TACTTGATAT	7680
TGTCAATTAT	GATGGTTCAA	TCTCAACAAC	AAGATTTAAG	AATAATAACA	TAAGCTTTGA	7740
TCAACCATAT	GCTGCACTAT	ACCCATCTGT	TGGACCAGGG	ATATACTACA	AAGGCAAAAT	7800
AATATTTCTC	GGGTATGGAG	GTCTTGAACA	TCCAATAAAT	GAGAATGTAA	TCTGCAACAC	7860
AACTGGGTGC	CCCGGGAAAA	CACAGAGAGA	CTGTAATCAA	GCGTCTCATA	GTCCATGGTT	7920
TTCAGATAGG	AGGATGGTCA	ACTCCATCAT	TGTTGTTGAC	AAAGGCTTAA	ACTCAATTCC	7980
AAAATTGAAA	GTATGGACGA	TATCTATGCG	ACAAAATTAC	TGGGGGTCAG	AAGGAAGGTT	8040
ACTTCTACTA	GGTAACAAGA	TCTATATATA	TACAAGATCT	ACAAGTTGGC	ATAGCAAGTT	8100
ACAATTAGGA	ATAATTGATA	TTACTGATTA	CAGTGATATA	AGGATAAAAT	GGACATGGCA	8160
TAATGTGCTA	TCAAGACCAG	GAAACAATGA	ATGTCCATGG	GGACATTCAT	GTCCAGATGG	8220
ATGTATAACA	GGAGTATATA	CTGATGCATA	TCCACTCAAT	CCCACAGGGA	GCATTGTGTC	8280
ATCTGTCATA	TTAGACTCAC	AAAAATCGAG	AGTGAACCCA	GTCATAACTT	ACTCAACAGC	8340
AACCGAAAGA	GTAAACGAGC	TGGCCATCCT	AAACAGAACA	CTCTCAGCTG	GATATACAAC	8400
AACAAGCTGC	ATTACACACT	ATAACAAAGG	ATATTGTTTT	CATATAGTAG	AAATAAATCA	8460
TAAAAGCTTA	AACACATTTC	AACCCATGTT	GTTCAAAACA	GAGATTCCAA	AAAGCTGCAG	8520

- 242 -

TTAATCATAA	<b>ȚTAACCATAA</b>	TATGCATCAA	TCTATCTATA	ATACAAGTAI	' ATGATAAGTA	8580
ATCAGCAATC	AGACAATAGA	CAAAAGGGAA	AAAAAATATA	CTTAGGAGCA	AAGCGTGCTC	8640
GGGAAATGGA	CACTGAATCT	AACAATGGCA	CTGTATCTGA	CATACTCTAT	CCTGAGTGTC	8700
ACCTTAACTC	TCCTATCGTT	AAAGGTAAAA	TAGCACAATT	ACACACTATT	ATGAGTCTAC	8760
CTCAGCCTTA	TGATATGGAT	GACGACTCAA	TACTAGTTAT	CACTAGACAG	AAAATAAAAC	8820
TTAATAAATT	GGATAAAAGA	CAACGATCTA	TTAGAAGATT	AAAATTAATA	TTAACTGAAA	8880
AAGTGAATGA	CTTAGGAAAA	TACACATTTA	TCAGATATCC	AGAAATGTCA	AAAGAAATGT	8940
TCAAATTATA	TATACCTGGT	ATTAACAGTA	AAGTGACTGA	ATTATTACTT	AAAGCAGATA	9000
GAACATATAG	TCAAATGACT	GATGGATTAA	GAGATCTATG	GATTAATGTG	CTATCAAAAT	9060
TAGCCTCAAA	AAATGATGGA	AGCAATTATG	ATCTTAATGA	AGAAATTAAT	AATATATCGA	9120
AAGTTCACAC	AACCTATAAA	TCAGATAAAT	GGTATAATCC	ATTCAAAACA	TGGTTTACTA	9180
TCAAGTATGA	TATGAGAAGA	TTACAAAAG	CTCGAAATGA	GATCACTTTT	AATGTTGGGA	9240
AGGATTATAA	CTTGTTAGAA	GACCAGAAGA	ATTTCTTATT	GATACATCCA	GAATTGGTTT	9300
TGATATTAGA	TAAACAAAAC	TATAATGGTT	ATCTAATTAC	TCCTGAATTA	GTATTGATGT	9360
ATTGTGACGT	AGTCGAAGGC	CGATGGAATA	TAAGTGCATG	TGCTAAGTTA	GATCCAAAAT	9420
TACAATCTAT	GTATCAGAAA	GGTAATAACC	TGTGGGAAGT	GATAGATAAA	TTGTTTCCAA	9480
TTATGGGAGA	AAAGACATTT	GATGTGATAT	CGTTATTAGA	ACCACTTGCA	TTATCCTTAA	9540
TTCAAACTCA	TGATCCTGTT	AAACAACTAA	GAGGAGCTTT	TTTAAATCAT	GTGTTATCCG	9600
AGATGGAATT	AATATTTGAA	TCTAGAGAAT	CGATTAAGGA	ATTTCTGAGT	GTAGATTACA	9660
TTGATAAAAT	TTTAGATATA	TTTAATAAGT	CTACAATAGA	TGAAATAGCA	GAGATTTTCT	9720
CTTTTTTAG	AACATTTGGG	CATCCTCCAT	TAGAAGCTAG	TATTGCAGCA	GAAAAGGTTA	9780
GAAAATATAT	GTATATTGGA	AAACAATTAA	AATTTGACAC	AAATAATAA	TGTCATGCTA	9840
TCTTCTGTAC	AATAATAATT	AACGGATATA	GAGAGAGGCA	TGGTGGACAG	TGGCCTCCTG	9900
TGACATTACC	TGATCATGCA	CACGAATTCA	TCATAAATGC	TTACGGTTCA	AACTCTGCGA	9960
TATCATATGA	AAATGCTGTT	GATTATTACC	AGAGCTTTAT	AGGAATAAAA	TTCAATAAAT	10020
TCATAGAGCC	TCAGTTAGAT	GAGGATTTGA	CAATTTATAT	gaaagataaa	GCATTATCTC	10080

- 243 -

	CAAAAAAATC	AAATTGGGAC	ACAGTTTATC	CTGCATCTAA	TTTACTGTAC	CGTACTAACG	10140
	CATCCAACGA	ATCACGAAGA	TTAGTTGAAG	TATTTATAGC	AGATAGTAAA	TTTGATCCTC	10200
•	ATCAGATATT	GGATTATGTA	GAATCTGGGG	ACTGGTTAGA	TGATCCAGAA	TTTATATTT	10260
	CTTATAGTCT	TAAAGAAAA	GAGATCAAAC	aggaaggtag	ACTCTTTGCA	AAAATGACAT	10320
	acaaaatgag	AGCTACACAA	GTTTTATCAG	AGACACTACT	TGCAAATAAC	ATAGGAAAAT	10380
	TCTTTCAAGA	AAATGGGATG	GTGAAGGGAG	AGATTGAATT	ACTTAAGAGA	TTAACAACCA	10440
	TATCAATATC	AGGAGTTCCA	CGGTATAATG	AAGTGTACAA	TAATTCTAAA	AGCCATACAG	10500
	ATGACCTTAA	AACCTACAAT	AAAATAAGTA	ATCTTAATTT	GTCTTCTAAT	CAGAAATCAA	10560
•	AGAAATTTGA	ATTCAAGTCA	ACGGATATCT	ACAATGATGG	ATACGAGACT	GTGAGCTGTT	10620
•	TCCTAACAAC	AGATCTCAAA	AAATACTGTC	TTAATTGGAG	ATATGAATCA	ACAGCTCTAT	10680
•	TTGGAGAAAC	TTGCAACCAA	ATATTTGGAT	TAAATAAATT	GTTTAATTGG	TTACACCCTC	10740
(	GTCTTGAAGG	AAGTACAATC	TATGTAGGTG	ATCCTTACTG	TCCTCCATCA	GATAAAGAAC	10800
1	ATATATCATT	AGAGGATCAC	CCTGATTCTG	GTTTTTACGT	TCATAACCCA	AGAGGGGGTA	10860
•	<b>PAGAAGGATT</b>	TTGTCAAAAA	TTATGGACAC	TCATATCTAT	AAGTGCAATA	CATCTAGCAG	10920
(	CTGTTAGAAT	AGGCGTGAGG	GTGACTGCAA	TGGTTCAAGG	AGACAATCAA	GCTATAGCTG	10980
7	PAACCACAAG	AGTACCCAAC	AATTATGACT	ACAGAGTTAA	GAAGGAGATA	GTTTATAAAG	11040
1	ATGTAGTGAG	ATTTTTTGAT	TCATTAAGAG	aagtgatgga	TGATCTAGGT	CATGAACTTA	11100
1	AATTAAATGA	AACGATTATA	AGTAGCAAGA	TGTTCATATA	TAGCAAAAGA	ATCTATTATG	11160
1	ATGGGAGAAT	TCTTCCTCAA	GCTCTAAAAG	CATTATCTAG	ATGTGTCTTC	TGGTCAGAGA	11220
(	CAGTAATAGA	CGAAACAAGA	TCAGCATCTT	CAAATTTGGC	AACATCATTT	GCAAAAGCAA	11280
•	rtgagaatgg	TTATTCACCT	GTTCTAGGAT	ATGCATGCTC	AATTTTTAAG	AACATTCAAC	11340
1	ACTATATAT	TGCCCTTGGG	ATGAATATCA	ATCCAACTAT	AACACAGAAT	ATCAGAGATC	11400
2	AGTATTTTAG	GAATCCAAAT	TGGATGCAAT	ATGCCTCTTT	AATACCTGCT	AGTGTTGGGG	11460
(	SATTCAATTA	CATGGCCATG	TCAAGATGTT	TTGTAAGGAA	TATTGGTGAT	CCATCAGTTG	11520
(	CCGCATTGGC	TGATATTAAA	AGATTTATTA	AGGCGAATCT	ATTAGACCGA	AGTGTTCTTT	11580
Z	TAGGATTAT	GAATCAAGAA	CCAGGTGAGT	CATCTTTTTT	GGACTGGGCT	TCAGATCCAT	11640

- 244 -

ATTCATGCAA	TTTACCACAA	TCTCAAAATA	TAACCACCAT	GATAAAAAAT	ATAACAGCAA	11700
GGAATGTATT	ACAAGATTCA	CCAAATCCAT	TATTATCTGG	ATTATTCACA	AATACAATGA	11760
TAGAAGAAGA	TGAAGAATTA	GCTGAGTTCC	TGATGGACAG	GAAGGTAATT	CTCCCTAGAG	11820
TTGCACATGA	TATTCTAGAT	AATTCTCTCA	CAGGAATTAG	AAATGCCATA	GCTGGAATGT	11880
TAGATACGAC	AAAATCACTA	ATTCGGGTTG	GCATAAATAG	AGGAGGACTG	ACATATAGTT	11940
TGTTGAGGAA	AATCAGTAAT	TACGATCTAG	TACAATATGA	AACACTAAGT	AGGACTTTGC	12000
GACTAATTGT	AAGTGATAAA	ATCAAGTATG	AAGATATGTG	TTCGGTAGAC	CTTGCCATAG	12060
CATTGCGACA	AAAGATGTGG	ATTCATTTAT	CAGGAGGAAG	GATGATAAGT	GGACTTGAAA	12120
CGCCTGACCC	ATTAGAATTA	CTATCTGGGG	TAGTAATAAC	AGGATCAGAA	CATTGTAAAA	12180
TATGTTATTC	TTCAGATGGC	ACAAACCCAT	ATACTTGGAT	GTATTTACCC	GGTAATATCA	12240
AAATAGGATC	AGCAGAAACA	GGTATATCGT	CATTAAGAGT	TCCTTATTTT	GGATCAGTCA	12300
CTGATGAAAG	ATCTGAAGCA	CAATTAGGAT	ATATCAAGAA	TCTTAGTAAA	CCTGCAAAAG	12360
CCGCAATAAG	AATAGCAATG	ATATATACAT	GGGCATTTGG	TAATGATGAG	ATATCTTGGA	12420
TGGAAGCCTC	ACAGATAGCA	CAAACACGTG	CAAATTTTAC	ACTAGATAGT	CTCAAAATTT	12480
TAACACCGGT	AGCTACATCA	ACAAATTTAT	CACACAGATT	AAAGGATACT	GCAACTCAGA	12540
TGAAATTCTC	CAGTACATCA	TTGATCAGAG	TCAGCAGATT	CATAACAATG	TCCAATGATA	12600
ACATGTCTAT	CAAAGAAGCT	AATGAAACCA	AAGATACTAA	TCTTATTTAT	CAACAAATAA	12660
TGTTAACAGG	ATTAAGTGTT	TTCGAATATT	TATTTAGATT	AAAAGAAACC	ACAGGACACA	12720
ACCCTATAGT	TATGCATCTG	CACATAGAAG	ATGAGTGTTG	TATTAAAGAA	AGTTTTAATG	12780
ATGAACATAT	TAATCCAGAG	TCTACATTAG	AATTAATTCG	ATATCCTGAA	AGTAATGAAT	12840
TTATTTATGA	TAAAGACCCA	CTCAAAGATG	TGGACTTATC	AAAACTTATG	GTTATTAAAG	12900
ACCATTCTTA	CACAATTGAT	ATGAATTATT	GGGATGATAC	TGACATCATA	CATGCAATTT	12960
CANTATGTAC	TGCAATTACA	ATAGCAGATA	CTATGTCACA	ATTAGATCGA	GATAATTTAA	13020
AAGAGATAAT	AGTTATTGCA	AATGATGATG	ATATTAATAG	CTTAATCACT	GAATTTTTGA	13080
CTCTTGACAT	ACTTGTATTT	CTCAAGACAT	TTGGTGGATT	ATTAGTAAAT	CAATTTGCAT	13140
ACACTCTTTA	TAGTCTAAAA	atagaaggta	GGGATCTCAT	TTGGGATTAT	ATAATGAGAA	13200

- 245 -

CACTGAGAGA	TACTTCCCAT	TCAATATTAA	AAGTATTATC	TAATGCATTA	TCTCATCCTA	13260
AAGTATTCAA	GAGGTTCTGG	GATTGTGGAG	TTTTAAACCC	TATTTATGGT	CCTAATACTG	13320
CTAGTCAAGA	CCAGATAAAA	CTTGCCCTAT	CTATATGTGA	ATATTCACTA	GATCTATTTA	13380
TGAGAGAATG	GTTGAATGGT	GTATCACTTG	AAATATACAT	TTGTGACAGC	GATATGGAAG	13440
TTGCAAATGA	TAGGAAACAA	GCCTTTATTT	CTAGACACCT	TTCATTTGTT	TGTTGTTTAG	13500
CAGAAATTGC	ATCTTTCGGA	CCTAACCTGT	TAAACTTAAC	ATACTTGGAG	AGACTTGATC	13560
TATTGAAACA	ATATCTTGAA	TTAAATATTA	AAGAAGACCC	TACTCTTAAA	TATGTACAAA	13620
TATCTGGATT	ATTAATTAAA	TCGTTCCCAT	CAACTGTAAC	ATACGTAAGA	AAGACTGCAA	13680
TCAAATATCT	AAGGATTCGC	GGTATTAGTC	CACCTGAGGT	AATTGATGAT	TGGGATCCGG	13740
TAGAAGATGA	AAATATGCTG	GATAACATTG	TCAAAACTAT	AAATGATAAC	TGTAATAAAG	13800
ATAATAAAGG	GAATAAAATT	AACAATTTCT	GGGGACTAGC	ACTTAAGAAC	TATCAAGTCC	13860
TTAAAATCAG	ATCTATAACA	AGTGATTCTG	ATGATAATGA	TAGACTAGAT	GCTAATACAA	13920
GTGGTTTGAC	ACTTCCTCAA	GGAGGGAATT	ATCTATCGCA	TCAATTGAGA	TTATTCGGAA	13980
TCAACAGCAC	TAGTTGTCTG	AAAGCTCTTG	AGTTATCACA	AATTTTAATG	AAGGAAGTCA	14040
ATAAAGACAA	GGACAGGCTC	TTCCTGGGAG	AAGGAGCAGG	AGCTATGCTA	GCATGTTATG	14100
ATGCCACATT	AGGACCTGCA	GTTAATTATT	ATAATTCAGG	TTTGAATATA	ACAGATGTAA	14160
TTGGTCAACG	AGAATTGAAA	ATATTTCCTT	CAGAGGTATC	ATTAGTAGGT	AAAAAATTAG	14220
GAAATGTGAC	ACAGATTCTT	AACAGGGTAA	AAGTACTGTT	CAATGGGAAT	CCTAATTCAA	14280
CATGGATAGG	AAATATGGAA	TGTGAGAGCT	TAATATGGAG	TGAATTAAAT	GATAAGTCCA	14340
TTGGATTAGT	ACATTGTGAT	ATGGAAGGAG	CTATCGGTAA	ATCAGAAGAA	ACTGTTCTAC	14400
ATGAACATTA	TAGTGTTATA	AGAATTACAT	ACTTGATTGG	GGATGATGAT	GTTGTTTTAG	14460
TTTCCAAAAT	TATACCTACA	ATCACTCCGA	ATTGGTCTAG	AATACTTTAT	CTATATAAAT	14520
TATATTGGAA	agatgtaagt	ATAATATCAC	TCAAAACTTC	TAATCCTGCA	TCAACAGAAT	14580
TATATCTAAT	TTCGAAAGAT	GCATATTGTA	CTATAATGGA	ACCTAGTGAA	ATTGTTTTAT	14640
CAAAACTTAA	AAGATTGTCA	CTCTTGGAAG	AAAATAATCT	ATTAAAATGG	ATCATTTAT	14700
CAAAGAAGAG	GAATAATGAA	TGGTTACATC	ATGAAATCAA	agaaggagaa	AGAGATTATG	14760

- 246 -

GAATCATGAG	ACCATATCAT	ATGGCACTAC	AAATCTTTGG	ATTTCAAATC	AATTTAAATC	14820
ATCTGGCGAA	AGAATTTTTA	TCAACCCCAG	ATCTGACTAA	TATCAACAAT	ATAATCCAAA	14880
GTTTTCAGCG	AACAATAAAG	GATGTTTTAT	TTGAATGGAT	TAATATAACT	CATGATGATA	14940
AGAGACATAA	ATTAGGCGGA	AGATATAACA	TATTCCCACT	GAAAAATAAG	GGAAAGTTAA	15000
GACTGCTATC	GAGAAGACTA	GTATTAAGTT	GGATTTCATT	ATCATTATCG	ACTCGATTAC	15060
TTACAGGTCG	CTTTCCTGAT	GAAAAATTTG	AACATAGAGC	ACAGACTGGA	TATGTATCAT	15120
TAGCTGATAC	TGATTTAGAA	TCATTAAAGT	TATTGTCGAA	AAACATCATT	AAGAATTACA	15180
GAGAGTGTAT	AGGATCAATA	TCATATTGGT	TTCTAACCAA	agaagttaaa	ATACTTATGA	15240
AATTGATTGG	TGGTGCTAAA	TTATTAGGAA	TTCCCAGACA	ATATAAAGAA	CCCGAAGACC	15300
AGTTATTAGA	AAACTACAAT	CAACATGATG	AATTTGATAT	CGATTAAAAC	ATAAATACAA	15360
rgaagatata	TCCTAACCTT	TATCTTTAAG	CCTAGGAATA	GACAAAAAGT	AAGAAAACA	15420
rgtaatatat	ATATACCAAA	CAGAGTTCTT	CTCTTGTTTG	GT		15462

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2233 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Asp Thr Glu Ser Asn Asn Gly Thr Val Ser Asp Ile Leu Tyr Pro 1 10 15

Glu Cys His Leu Asn Ser Pro Ile Val Lys Gly Lys Ile Ala Gln Leu 20 25 30

His Thr Ile Met Ser Leu Pro Gln Pro Tyr Asp Met Asp Asp Asp Ser 35 40 45

Ile Leu Val Ile Thr Arg Gln Lys Ile Lys Leu Asn Lys Leu Asp Lys 50 55 60

- 247 -

Arg 65	Gln	Arg	Ser	Ile	Arg 70	Arg	Leu	Lys	Leu	Ile 75	Leu	Thr	Glu	. Lys	Val 80
Asn	Asp	Leu	Gly	Lys 85	Tyr	Thr	Phe	Ile	Arg 90	Tyr	Pro	Glu	Met	Ser 95	Lys
Glu	Met	Phe	Lys 100		Tyr	Ile	Pro	Gly 105		Asn	Ser	Lys	Val	Thr	Glu
Leu	Leu	Leu 115		Ala	Asp	Arg	Thr 120	Tyr	Ser	Gln	Met	Thr 125		Gly	Leu
Arg	Asp 130	Leu	Trp	Ile	Asn	Val 135	Leu	Ser	Lys	Leu	Ala 140		Lys	Asn	Двр
Gly 145	Ser	Asn	Tyr	Asp	Leu 150		Glu	Glu	Ile	Asn 155	Asn	Ile	Ser	Lys	Val 160
His	Thr	Thr	Tyr	Lув 165	Ser	Авр	Lys	Trp	Tyr 170	Asn	Pro	Phe	Lys	Thr 175	Trp
Phe	Thr	Ile	Lys 180		Asp	Met	Arg	Arg 185	Leu	Gln	Lys	Ala	Arg 190	Asn	Glu
Ile	Thr	Phe 195	Asn	Val	Gly	Lys	Авр 200	Tyr	Asn	Leu	Leu	Glu 205	Asp	Gln	Lys
Asn	Phe 210	Leu	Leu	Ile	His	Pro 215	Glu	Leu	Val	Leu	Ile 220	Leu	Asp	Lys	Gln
Asn 225	Tyr	Asn	Gly	Tyr	Leu 230	Ile	Thr	Pro	Glu	Leu 235	Val	Leu	Met	Tyr	Сув 240
yab	Val	Val	Glu	Gly 245	Arg	Trp	Asn	Ile	Ser 250	Ala	Сув	Ala	Lys	Leu 255	Asp
Pro	Lya	Leu	Gln 260	Ser	Met	Tyr	Gln	<b>Lys</b> 265	Gly	Asn	Asn	Leu	Trp 270	Glu	Val
Ile	Asp	Lys 275	Leu	Phe	Pro	Ile	Met 280	Gly	Glu	Lys	Thr	Phe 285	Asp	Val	Ile
Ser	<b>Leu</b> 290	Leu	Glu	Pro	Leu	Ala 295	Leu	Ser	Leu	Ile	Gln 300	Thr	His	Asp	Pro
Val 305	Lys	Gln	Leu	Arg	Gly 310	Ala	Phe	Leu	Asn	His 315	Val	Leu	Ser	Glu	Met 320
Glu	Leu	Ile	Phe	Glu 325	Ser	Arg	Glu	Ser	Ile 330	Lys	Glu	Phe	Leu	Ser 335	Val
qaA	Tyr	Ile	Asp	Lys	Ile	Leu	Asp	Ile	Phe	Asn	Lys	Ser	Thr	Ile	Asp

- 248 -

			340					345					350		
Glu	Ile	Ala 355	Glu	Ile	Phe	Ser	Phe 360	Phe	Arg	Thr	Phe	Gly 365	His	Pro	Pro
Leu	Glu 370	Ala	Ser	Ile	Ala	Ala 375	Glu	Lys	Val	Arg	180	Tyr	Met	Tyr	Ile
Gly 385	Lys	Gln	Leu	Lys	Phe 390	Asp	Thr	Ile	Asn	Lув 395	Сув	His	Ala	Ile	Phe 400
Сув	Thr	Ile	Ile	11e 405	Asn	Gly	Tyr	Arg.	Glu 410	Arg	His	Gly	Gly	Gln 415	Trp
Pro	Pro	Val	Thr 420	Leu	Pro	Asp	His	Ala 425	His	Glu	Phe	Ile	11e 430	Asn	Ala
Tyr	Gly	Ser 435	Asn	Ser	Ala	Ile	Ser 440	Tyr	Glu	Asn	Ala	Val 445	Asp	Tyr	Тут
Gln	Ser 450	Phe	Ile	Gly	Ile	Lys 455	Phe	Asn	Lys	Phe	Ile 460	Glu	Pro	Gln	Leu
Asp 465	Glu	Asp	Leu	Thr	11e 470	Tyr	Met	Lys	Asp	Lys 475	Ala	Leu	Ser	Pro	Lys 480
Lys	Ser	Asn	Trp	Авр 485	Thr	Val	Tyr	Pro	Ala 490	Ser	Asn	Leu	Leu	Tyr 495	Arg
Thr	Asn	Ala	Ser 500	Asn	Glu	Ser	Arg	Arg 505	Leu	Val	Glu	Val	Phe 510	Ile	Ala
	Ser	515					520				7	525			
qaA	Trp 530	Leu	Asp	Asp	Pro	Glu 535	Phe	Asn	Ile	Ser	Tyr 540	Ser	Leu	Lys	Glu
Lys 545	Glu	Ile	Lys	Gln	Glu 550	Gly	Arg	Leu	Phe	Ala 555	Lys	Met	Thr	Tyr	Lys 560
Met	Arg	Ala	Thr	Gln 565	Val	Leu	Ser	Glu	Thr 570	Leu	Leu	Ala	Asn	Asn 575	Ile
Gly	Lys	Phe	Phe 580	Gln	Glu	Asn	Gly	Met 585	Val	Lys	Gly	Glu	Ile 590	Glu	Leu
	Lys	595					600					605			
Glu	Val 610	Tyr	Asn	Asn	Ser	Lys 615	Ser	His	Thr	Asp	Asp 620	Leu	Lys	Thr	Tyr

- 249 -

Asn 625	Lys	Ile	Ser	Asn	Leu 630	Asn	Leu	Ser	Ser	Asn 635	Gln	Lys	Ser	Lys	Lys 640
Phe	Glu	Phe	Lys	8er 645	Thr	Asp	Ile	Tyr	Asn 650	Asp	Gly	Tyr	Glu	Thr 655	Val
Ser	Сув	Phe	Leu 660	Thr	Thr	Авр	Leu	Lys 665	Lys	Tyr	Сув	Leu	Asn 670	Trp	Arg
Tyr	Glu	Ser 675	Thr	Ala	Leu	Phe	Gly 680	Glu	Thr	Сув	Asn	Gln 685	Ile	Phe	Gly
Leu	Asn 690	Lys	Leu	Phe	Asn	Trp 695	Leu	His	Pro	Arg	Leu 700	Glu	Gly	Ser	Thr
11e 705	Tyr	Val	Gly	Asp	Pro 710	Tyr	Сув	Pro	Pro	Ser 715	Asp	Lys	Glu	His	11e 720
Ser	Leu	Glu	Asp	His 725	Pro	qaA	Ser	Gly	Phe 730	Tyr	Val	His	Asn	Pro 735	Arg
Gly	Gly	Ile	Glu 740	Gly	Phe	Сув	Gln	Lys 745	Leu	Trp	Thr	Leu	Ile 750	Ser	Ile
Ser	Ala	11e 755	His	Leu	Ala	Ala	Val 760	Arg	Ile	Gly	Val	Arg 765	Val	Thr	Ala
Met	Val 770	G1n	Gly	увр	Asn	Gln 775	Ala	Ile	Ala	Val	Thr 780	Thr	Arg	Val	Pro
Asn 785	Asn	Tyr	Asp	Tyr	Arg 790	Val	Lys	Lys	Glu	Ile 795	Val	Tyr	Lys	Авр	Val 800
	Arg			805					810					815	
Glu	Leu	Lys	Leu 820	Asn	Glu	Thr	Ile	11e 825	Ser	Ser	Lys	Met	Phe 830	Ile	Tyr
Ser	Lys	Arg 835	Ile	Tyr	Tyr	Asp	Gly 840	Arg	Ile	Leu	Pro	Gln 845	Ala	Leu	Lys
Ala	Leu 850	Ser	Arg	Сув	Val	Phe 855	Trp	Ser	Glu	Thr	Val 860	Ile	Asp	Glu	Thr
Arg 865	Ser	Ala	Ser	Ser	<b>Asn</b> 870	Leu	Ala	Thr	Ser	Phe 875	Ala	Lys	Ala	Ile	Glu 880
Asn	Gly	Tyr	Ser	Pro 885	Val	Leu	Gly	Tyr	<b>Ala</b> 890	Сув	Ser	Ile	Phe	Lув 895	Asn

- 250 -

Ile Gln Gln Leu Tyr Ile Ala Leu Gly Met Asn Ile Asn Pro Thr Ile Thr Gln Asn Ile Arg Asp Gln Tyr Phe Arg Asn Pro Asn Trp Met Gln 920 Tyr Ala Ser Leu Ile Pro Ala Ser Val Gly Gly Phe Asn Tyr Met Ala 935 Met Ser Arg Cys Phe Val Arg Asn Ile Gly Asp Pro Ser Val Ala Ala 945 955 Leu Ala Asp Ile Lys Arg Phe Ile Lys Ala Asn Leu Leu Asp Arg Ser 970 Val Leu Tyr Arg Ile Met Asn Gln Glu Pro Gly Glu Ser Ser Phe Leu 985 Asp Trp Ala Ser Asp Pro Tyr Ser Cys Asn Leu Pro Gln Ser Gln Asn 995 1000 Ile Thr Thr Met Ile Lys Asn Ile Thr Ala Arg Asn Val Leu Gln Asp 1015 Ser Pro Asn Pro Leu Leu Ser Gly Leu Phe Thr Asn Thr Met Ile Glu 1025 1030 1035 Glu Asp Glu Glu Leu Ala Glu Phe Leu Met Asp Arg Lys Val Ile Leu 1045 Pro Arg Val Ala His Asp Ile Leu Asp Asn Ser Leu Thr Gly Ile Arg 1060 Asn Ala Ile Ala Gly Met Leu Asp Thr Thr Lys Ser Leu Ile Arg Val 1075 1080 Gly Ile Asn Arg Gly Gly Leu Thr Tyr Ser Leu Leu Arg Lys Ile Ser Asn Tyr Asp Leu Val Gln Tyr Glu Thr Leu Ser Arg Thr Leu Arg Leu 1105 1110 Ile Val Ser Asp Lys Ile Lys Tyr Glu Asp Met Cys Ser Val Asp Leu 1130 Ala Ile Ala Leu Arg Gln Lys Met Trp Ile His Leu Ser Gly Gly Arg 1140 1150 Met Ile Ser Gly Leu Glu Thr Pro Asp Pro Leu Glu Leu Ser Gly 1160

Val Val Ile Thr Gly Ser Glu His Cys Lys Ile Cys Tyr Ser Ser Asp

- 251 -

	1170	ı				1175	;				1180	)			
Gly 1185		Asn	Pro	Tyr	Thr 1190		Met	Tyr	Leu	Pro 1195		Asn	Ile	Lys	Ile 1200
Gly	Ser	Ala	Glu	Thr 1205		Ile	Ser	Ser	Leu 1210	Arg	Val	Pro	Tyr	Phe 1215	
Ser	Val	Thr	Asp 1220		Arg	Ser	Glu	Ala 1225		Leu	Gly	Tyr	Ile 1230		Asn
Leu	Ser	Lys 1235		Ala	Lys	Ala	Ala 1240		Arg	Ile	Ala	Met 1245		Tyr	Thr
Trp	Ala 1250		Gly	Asn	Авр	Glu 1255		Ser	Trp	Met	Glu 1260		Ser	Gln	Ile
Ala 1265		Thr	Arg	Ala	Asn 1270		Thr	Leu	Авр	Ser 1275		Lys	Ile	Leu	Thr 1280
Pro	Val	Ala	Thr	Ser 128		Asn	Leu	Ser	His 129	Arg )	Leu	Lys	Asp	Thr 1295	
Thr	Gln	Met	Lys 130		Ser	Ser	Thr	Ser 1305		Ile	Arg	Val	Ser 131		Phe
_						_				_	_		_		
Ile	Thr	Met 131		Asn	Asp	Asn	1320		Ile	Lys	Glu	1325		GIU	Thr
		131:	5				1320 Gln	)		Lys Met		1325 Thr	5		
Lys	Asp 1330 Phe	131! Thr	Asn	Leu	Ile	Tyr 133! Arg	1320 Gln 5	Gln	Ile	Met	Leu 1340 Thr	1325 Thr	Gly	Leu	
Lys Val 134	Asp 1330 Phe	Thr	Asn Tyr	Leu	Ile Phe 1350	Tyr 133! Arg	1320 Gln 5 Leu	Gln Lys	Ile Glu	Met Thr 1355	Leu 1340 Thr	Thr	Gly His	Leu Asn	Ser Pro 1360 Ser
Lys Val 134	Asp 1330 Phe 5	Thr O Glu Met	Asn Tyr His	Leu Leu Leu 1369	Phe 1350 His	Tyr 133: Arg	Gln 5 Leu Glu	Gln Lys Asp	Ile Glu Glu 137	Thr 135! Cys	Leu 1340 Thr Cys	Thr O Gly	Gly His Lys	Leu Asn Glu 1379	Ser Pro 1360 Ser
Lys Val 134: Ile	Asp 1330 Phe 5 Val	Thr Glu Met	Asn Tyr His Glu 138 Ser	Leu Leu 1369 His	Phe 1350 His 5	Tyr 133: Arg D	Gln Leu Glu Pro	Gln Lys Asp Glu 138:	Glu Glu 137	Met Thr 1359 Cys O Thr	Leu 1340 Thr Cys	Thr Gly Ile	Gly His Lys Leu 139	Leu Asn Glu 137! Ile	Pro 1360 Ser
Lys Val 134: Ile Phe	Asp 1330 Phe 5 Val Asn	Thr Glu Met Asp Glu 139	Asn Tyr His Glu 138 Ser	Leu Leu 136: His O	Phe 1356 His 5	Tyr 1339 Arg Ile Asn	Gln  Leu  Glu  Pro  11e 1400	Gln Lys Asp Glu 138:	Glu Glu 1370 Ser 5	Thr 1359 Cys Thr	Leu 1340 Thr Cys Leu Asp	Thr Gly Ile Glu Pro 1409	Gly His Lys Leu 1390	Asn Glu 1379 Ile D	Pro 1360 Ser Arg
Lys Val 134 Ile Phe Tyr	Asp 1330 Phe 5 Val Asn Pro Asp 1410	Thr Glu Met Asp Glu 139 Leu 0	Asn Tyr His Glu 138 Ser 5	Leu Leu 136: His O Asn	Phe 1356 His 5 Ile Glu Leu	Tyr 1339 Arg Ile Asn Phe Met 141	Gln  Glu  Glu  Pro  Ile 1400  Val	Gln Lys Asp Glu 138:	Glu Glu 137 Ser 5	Thr 135: Cys Thr Lys	Leu 1340 Thr Cys Leu Asp His 1423	Thr Gly Ile Glu Pro 1409	Gly His Lys Leu 1390 Leu 5	Leu Asn Glu 1379 Ile D	Pro 1360 Ser Arg

- Asn Leu Lys Glu Ile Ile Val Ile Ala Asn Asp Asp Asp Ile Asn Ser 1460 1465 1470
- Leu Ile Thr Glu Phe Leu Thr Leu Asp Ile Leu Val Phe Leu Lys Thr 1475 1480 1485
- Phe Gly Gly Leu Leu Val Asn Gln Phe Ala Tyr Thr Leu Tyr Ser Leu 1490 1495 1500
- Lys Ile Glu Gly Arg Asp Leu Ile Trp Asp Tyr Ile Met Arg Thr Leu 1505 1510 1515 1520
- Arg Asp Thr Ser His Ser Ile Leu Lys Val Leu Ser Asn Ala Leu Ser 1525 1530 1535
- His Pro Lys Val Phe Lys Arg Phe Trp Asp Cys Gly Val Leu Asn Pro 1540 1545 1550
- Ile Tyr Gly Pro Asn Thr Ala Ser Gln Asp Gln Ile Lys Leu Ala Leu 1555 1560 1565
- Ser Ile Cys Glu Tyr Ser Leu Asp Leu Phe Met Arg Glu Trp Leu Asn 1570 1575 1580
- Gly Val Ser Leu Glu Ile Tyr Ile Cys Asp Ser Asp Met Glu Val Ala 1585 1590 1595 1600
- Asn Asp Arg Lys Gln Ala Phe Ile Ser Arg His Leu Ser Phe Val Cys 1605 1610 1615
- Cys Leu Ala Glu Ile Ala Ser Phe Gly Pro Asn Leu Leu Asn Leu Thr 1620 1625 1630
- Tyr Leu Glu Arg Leu Asp Leu Leu Lys Gln Tyr Leu Glu Leu Asn Ile 1635 1640 1645
- Lys Glu Asp Pro Thr Leu Lys Tyr Val Gln Ile Ser Gly Leu Leu Ile 1650 1655 1660
- Lys Ser Phe Pro Ser Thr Val Thr Tyr Val Arg Lys Thr Ala Ile Lys 1665 1670 1675 1680
- Tyr Leu Arg Ile Arg Gly Ile Ser Pro Pro Glu Val Ile Asp Asp Trp 1685 1690 1695
- Asp Pro Val Glu Asp Glu Asn Met Leu Asp Asn Ile Val Lys Thr Ile 1700 1705 1710
- Asn Asp Asn Cys Asn Lys Asp Asn Lys Gly Asn Lys Ile Asn Asn Phe 1715 1720 1725

- 253 -

Trp	Gly	Leu	Ala	Leu	Lys	Asn	Tyr	Gln	Val	Leu	Lys	Ile	Arg	Ser	Ile
	1730	)				173	5				1740	0			

- Thr Ser Asp Ser Asp Asp Asp Asp Arg Leu Asp Ala Asn Thr Ser Gly 1745 1750 1755 1766
- Leu Thr Leu Pro Gln Gly Gly Asn Tyr Leu Ser His Gln Leu Arg Leu 1765 1770 1775
- Phe Gly Ile Asn Ser Thr Ser Cys Leu Lys Ala Leu Glu Leu Ser Gln 1780 1785 1790
- Ile Leu Met Lys Glu Val Asn Lys Asp Lys Asp Arg Leu Phe Leu Gly 1795 1800 1805
- Glu Gly Ala Gly Ala Met Leu Ala Cys Tyr Asp Ala Thr Leu Gly Pro 1810 1815 1820
- Ala Val Asn Tyr Tyr Asn Ser Gly Leu Asn Ile Thr Asp Val Ile Gly 1825 1830 1835 1840
- Gln Arg Glu Leu Lys Ile Phe Pro Ser Glu Val Ser Leu Val Gly Lys 1845 1850 1855
- Lys Leu Gly Asn Val Thr Gln Ile Leu Asn Arg Val Lys Val Leu Phe 1860 1865 1870
- Asn Gly Asn Pro Asn Ser Thr Trp Ile Gly Asn Met Glu Cys Glu Ser 1875 1880 1885
- Leu Ile Trp Ser Glu Leu Asn Asp Lys Ser Ile Gly Leu Val His Cys 1890 1895 1900
- Asp Met Glu Gly Ala Ile Gly Lys Ser Glu Glu Thr Val Leu His Glu 1905 1910 1915 1920
- His Tyr Ser Val Ile Arg Ile Thr Tyr Leu Ile Gly Asp Asp Val 1925 1930 1935
- Val Leu Val Ser Lys Ile Ile Pro Thr Ile Thr Pro Asn Trp Ser Arg 1940 1945 1950
- Ile Leu Tyr Leu Tyr Lys Leu Tyr Trp Lys Asp Val Ser Ile Ile Ser 1955 1960 1965
- Leu Lys Thr Ser Asn Pro Ala Ser Thr Glu Leu Tyr Leu Ile Ser Lys 1970 1975 1980
- Asp Ala Tyr Cys Thr Ile Met Glu Pro Ser Glu Ile Val Leu Ser Lys 1985 1990 1995 2000
- Leu Lys Arg Leu Ser Leu Leu Glu Glu Asn Asn Leu Leu Lys Trp Ile

- 254 -

2005

2010

2015

- Ile Leu Ser Lys Lys Arg Asn Asn Glu Trp Leu His His Glu Ile Lys 2020 2025 2030
- Glu Gly Glu Arg Asp Tyr Gly Ile Met Arg Pro Tyr His Met Ala Leu 2035 2040 2045
- Gln Ile Phe Gly Phe Gln Ile Asn Leu Asn His Leu Ala Lys Glu Phe 2050 2055 2060
- Leu Ser Thr Pro Asp Leu Thr Asn Ile Asn Asn Ile Ile Gln Ser Phe 2065 2070 2075 2080
- Gln Arg Thr Ile Lys Asp Val Leu Phe Glu Trp Ile Asn Ile Thr His 2085 2090 2095
- Asp Asp Lys Arg His Lys Leu Gly Gly Arg Tyr Asn Ile Phe Pro Leu 2100 2105 2110
- Lys Asn Lys Gly Lys Leu Arg Leu Leu Ser Arg Arg Leu Val Leu Ser 2115 2120 2125
- Trp Ile Ser Leu Ser Leu Ser Thr Arg Leu Leu Thr Gly Arg Phe Pro 2130 2135 2140
- Asp Glu Lys Phe Glu His Arg Ala Gln Thr Gly Tyr Val Ser Leu Ala 2145 2150 2155 2160
- Asp Thr Asp Leu Glu Ser Leu Lys Leu Ser Lys Asn Ile Ile Lys 2165 2170 2175
- Asn Tyr Arg Glu Cys Ile Gly Ser Ile Ser Tyr Trp Phe Leu Thr Lys 2180 2185 2190
- Glu Val Lys Ile Leu Met Lys Leu Ile Gly Gly Ala Lys Leu Leu Gly 2195 2200 2205
- Ile Pro Arg Gln Tyr Lys Glu Pro Glu Asp Gln Leu Leu Glu Asn Tyr 2210 2215 2220
- Asn Gln His Asp Glu Phe Asp Ile Asp 2225 2230
- (2) INFORMATION FOR SEQ ID NO:19:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 15462 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear

- 255 -

## (ii) MOLECULE TYPE: RNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

60	AACTTAGGAT	ACTTAAAATT	TATAAATTTA	TGCTTGGTAA	AGAAGAAACT	ACCAAACAAG
120	ATGTTGAGCC	AATTTCAAAA	GGAACTCTAT	TCAAGAAAAG	GACTAGAAGG	TTAAGACATT
180	GGTGGAGCTA	AAAATCAGCC	AAAACATAAC	CGTAGGCAAG	ATTTAATGCA	TATTTGATAC
240	ATAACTGATG	TGGACCGACA	TATTCGCCCT	ACTGTCTCTA	ACAGAAAAT	TCATTCCTGG
300	AATGAGAAAC	TTCACTAGAT	TTCTATCTCA	GCTCTTCTAT	AATGACATTA	ATAATGAGAA
360	GCCAATCCAG	AATGGCTTAT	CTTTATTGTC	TTCTTGGTGT	AAGGCAGGG	AACATGCACA
420	TACATGATTG	GTATGTCATA	CAGATGCCAA	GGAAGTAATG	AACAACAAAT	AGCTCTACCT
480	GAGATGATAT	TAAGACGAGA	GATTTGTGGT	AAGTATGGAG	AAAACGGCAA	AGAAAGATCT
540	GAAACTATGT	TTATGATCAG	GTGACCTGGA	ATATTTGGAA	AACTGATTGG	ATGAAAAGAC
600	GGGTATCCAT	CCACACATTT	AAGACCTTGT	TCAACAATTG	CAGGAACAAT	TGCAGAACGG
660	ATCACTAGTA	GGTCAAAGCT	GGATAGTTCT	ATACAGATCT	AGCTCTTATA	CATGTTTAGG
720	GATGGAACAG	TTTCAGACAA	GATTGGAAGC	TTTTTCACCC	AAGAAAAGGC	TCTCAGGGTT
780	ATCATGCGGT	GATTGGGTCA	CAGTGGATCA	AGCGGTGACA	GCTGGTATTG	TGCAGGCAGG
840	ACCAGCAGAA	AACAATGAAT	AAACATTAAT	CTTATGGTTG	CTTGGTAACT	CTCAACAGAG
900	AGAGATGCAG	CAACTACATA	AAATTGTTGG	AAGAATATAC	AACCATAGAA	ATGACCTCAC
960	GCAGCTTTGA	GACCAGAATG	ATGGAATTGA	ACAATCAGAT	ATTCTTCAAT	GTCTCGCTTC
1020	CTGTATTTAT	TTTGATGGAA	GATTAAAAGC	GATATCAATA	TCTCAGACCA	CTCTATCCAC
1080	GGTGAGTTCG	TCCTATACAT	TCCTCAGAGA	TTCATCTGTA	ACGCGCTCCT	CAAAGGGACC
1140	GTACAAAATA	GGTGGCAGTT	ATGCAATGGG	ATATGGAGCT	CTATCCTGCC	CACCAGGCAA
1200	TTCCAGCTAG	CATTGATATG	CATATCTAGA	ACGGGAAGAT	ACAGTATGTG	GAGCCATGCA
1260	GATGAACTTG	AACACTGGAA	AAATGAGCTC	GCCGAAGCTC	AGCACGTGAT	GACAAGCAGT
1320	AACAGTTCAG	AAGGAACATA	AGAGACATAT	GAAAGCTTGA	CGAAGCTAAA	GAGTGACACA
1380	GATGAAGAGC	GATGGCAATA	CAGCCATAGA	ACAGGTGGAT	CCACAAACCG	AGACATCTTT

- 256 -

CAGAACAATT	CGAACATAGA	GCAGATCAAG	AACAAAATGG	AGAACCTCAA	TCATCCATAA	1440
TTCAATATGC	CTGGGCAGAA	GGAAATAGAA	GCGATGATCA	GACTGAGCAA	GCTACAGAAT	1500
CTGACAATAT	CAAGACCGAA	CAACAAAACA	TCAGAGACAG	ACTARACARG	AGACTCAACG	1560
ACAAGAAGAA	ACAAAGCAGT	CAACCACCCA	CTAATCCCAC	AAACAGAACA	AACCAGGACG	1620
AAATAGATGA	TCTGTTTAAC	GCATTTGGAA	GCAACTAATC	GAATCAACAT	TTTAATCTAA	1680
ATCAATAATA	AATAAGAAAA	ACTTAGGATT	AAAGAATCCT	ATCATACCGG	AATATAGGGT	1740
GGTAAATTTA	GAGTCTGCTT	GAAACTCAAT	CAATAGAGAG	TTGATGGAAA	GCGATGCTAA	1800
AAACTATCAA	ATCATGGATT	CTTGGGAAGA	GGAATCAAGA	GATAAATCAA	CTAATATCTC	1860
CTCGGCCCTC	AACATCATTG	AATTCATACT	CAGCACCGAC	CCCCAAGAAG	ACTTATCGGA	1920
AAACGACACA	ATCAACACAA	GAACCCAGCA	ACTCAGTGCC	ACCATCTGTC	AACCAGAAAT	1980
CAAACCAACA	GAAACAAGTG	agaaagatag	TGGATCAACT	GACAAAAATA	GACAGTCTGG	2040
GTCATCACAC	GAATGTACAA	CAGAAGCAAA	AGATAGAAAC	ATTGATCAGG	AAACTGTACA	2100
GAGAGGACCT	GGGAGAAGAA	GCAGCTCAGA	TAGTAGAGCT	GAGACTGTGG	TCTCTGGAGG	2160
AATCCCCAGA	AGCATCACAG	ATTCTAAAAA	TGGAACCCAA	AACACGGAGG	ATATTGATCT	2220
CAATGAAATT	AGAAAGATGG	ATAAGGACTC	TATTGAGGGG	AAAATGCGAC	AATCTGCAAA	2280
TGTTCCAAGC	GAGATATCAG	GAAGTGATGA	CATATTTACA	ACAGAACAAA	GTAGAAACAG	2340
TGATCATGGA	AGAAGCCTGG	AATCTATCAG	TACACCTGAT	ACAAGATCAA	TAAGTGTTGT	2400
TACTGCTGCA	ACACCAGATG	ATGAAGAAGA	AATACTAATG	AAAAATAGTA	GGACAAAGAA	2460
AAGTTCTTCA	ACACATCAAG	AAGATGACAA	AAGAATTAAA	AAAGGGGGAA	AAGGGAAAGA	2520
CTGGTTTAAG	AAATCAAAAG	ATACCGACAA	CCAGATACCA	ACATCAGACT	ACAGATCCAC	2580
atcaaaaggg	CAGAAGAAAA	TCTCAAAGAC	AACAACCACC	AACACCGACA	CAAAGGGGCA	2640
AACAGAAATA	CAGACAGAAT	CATCAGAAAC	ACAATCCTCA	TCATGGAATC	TCATCATCGA	2700
CAACAACACC	GACCGGAACG	AACAGACAAG	CACAACTCCT	CCAACAACAA	CTTCCAGATC	2760
AACTTATACA	AAAGAATCGA	TCCGAACAAA	CTCTGAATCC	AAACCCAAGA	CACAAAAGAC	2820
aaatggaaag	GAAAGGAAGG	ATACAGAAGA	GAGCAATCGA	TTTACAGAGA	GGGCAATTAC	2880
TCTATTGCAG	AATCTTGGTG	TAATTCAATC	CACATCAAAA	CTAGATTTAT	ATCAAGACAA	2940

- 257 -

ACGAGTTGTA TGTGTAGCAA ATGTACTAAA CAATGTAGAT ACTGCATCAA AGATAGATTT	3000
CCTGGCAGGA TTAGTCATAG GGGTTTCAAT GGACAACGAC ACAAAATTAA CACAGATACA	3060
ARATGARATG CTARACCTCA RAGCAGATCT RARGRARATG GRCGARTCAC ATAGRAGATT	3120
GATAGAAAAT CAAAGAGAAC AACTGTCATT GATCACGTCA CTAATTTCAA ATCTCAAAAT	3180
TATGACTGAG AGAGGAGGAA AGAAAGACCA AAATGAATCC AATGAGAGAG TATCCATGAT	3240
CAAAACAAAA TTGAAAGAAG AAAAGATCAA GAAGACCAGG TTTGACCCAC TTATGGAGGC	3300
ACAAGGCATT GACAAGAATA TACCCGATCT ATATCGACAT GCAGGAGATA CACTAGAGAA	3360
CGATGTACAA GTTAAATCAG AGATATTAAG TTCATACAAT GAGTCAAATG CAACAAGACT	3420
AATACCCAAA AAAGTGAGCA GTACAATGAG ATCACTAGTT GCAGTCATCA ACAACAGCAA	3480
TCTCTCACAA AGCACAAAAC AATCATACAT AAACGAACTC AAACGTTGCA AAAATGATGA	3540
AGAAGTATCT GAATTAATGG ACATGTTCAA TGAAGATGTC AACAATTGCC AATGATCCAA	3600
CAAAGAAACG ACACCGAACA AACAGACAAG AAACAACAGT AGATCAAAAC CTGTCAACAC	3660
ACACAAAATC AAGCAGAATG AAACAACAGA TATCAATCAA TATACAAATA AGAAAAACTT	3720
AGGATTAAAG AATAAATTAA TCCTTGTCCA AAATGAGTAT AACTAACTCT GCAATATACA	3780
CATTCCCAGA ATCATCATTC TCTGAAAATG GTCATATAGA ACCATTACCA CTCAAAGTCA	3840
ATGAACAGAG GAAAGCAGTA CCCCACATTA GAGTTGCCAA GATCGGAAAT CCACCAAAAC	3900
ACGGATCCCG GTATTTAGAT GTCTTCTTAC TCGGCTTCTT CGAGATGGAA CGAATCAAAG	3960
ACAAATACGG GAGTGTGAAT GATCTCGACA GTGACCCGAG TTACAAAGTT TGTGGCTCTG	4020
GATCATTACC AATCGGATTG GCTAAGTACA CTGGGAATGA CCAGGAATTG TTACAAGCCG	4080
CAACCAAACT GGATATAGAA GTGAGAAGAA CAGTCAAAGC GAAAGAGATG GTTGTTTACA	4140
CGGTACAAAA TATAAAACCA GAACTGTACC CATGGTCCAA TAGACTAAGA AAAGGAATGC	4200
TGTTCGATGC CAACAAAGTT GCTCTTGCTC CTCAATGTCT TCCACTAGAT AGGAGCATAA	4260
AATTTAGAGT AATCTTCGTG AATTGTACGG CAATTGGATC AATAACCTTG TTCAAAATTC	4320
CTAAGTCAAT GGCATCACTA TCTCTAACCA ACACAATATC AATCAATCTG CAGGTACACA	4380
TAAAAACAGG GGTTCAGACT GATTCTAAAG GGATAGTTCA AATTTTGGAT GAGAAAGGCG	4440
AAAAATCACT GAATTTCATG GTCCATCTCG GATTGATCAA AAGAAAAGTA GGCAGAATGT	4500

- 258 -

ACTCTGTTGA	ATACTGTAAA	CAGAAAATCG	agaaaatgag	ATTGATATTT	TCTTTAGGAC	4560
TAGTTGGAGG	AATCAGTCTT	CATGTCAATG	CAACTGGGTC	CATATCAAAA	ACACTAGCAA	4620
GTCAGCTGGT	ATTCAAAAGA	GAGATTTGTT	ATCCTTTAAT	GGATCTAAAT	CCGCATCTCA	4680
ATCTAGTTAT	CTGGGCTTCA	TCAGTAGAGA	TTACAAGAGT	GGATGCAATT	TTCCAACCTT	4740
CTTTACCTGG	CGAGTTCAGA	TACTATCCTA	ATATTATTGC	AAAAGGAGTT	GGGAAAATCA	4800
AACAATGGAA	CTAGTAATCT	CTATTTTAGT	CCGGACGTAT	CTATTAAGCC	GAAGCAAATA	4860
AAGGATAATC	AAAAACTTAG	GACAAAAGAG	GTCAATACCA	ACAACTATTA	GCAGTCACAC	4920
TCGCAAGAAT	AAGAGAGAAG	GGACCAAAAA	AGTCAAATAG	GAGAAATCAA	AACAAAAGGT	4980
ACAGAACACC	AGAACAACAA	AATCAAAACA	TCCAACTCAC	TCAAAACAAA	AATTCCAAAA	5040
GAGACCGGCA	ACACAACAAG	CACTGAACAC	AATGCCAACT	TCAATACTGC	TAATTATTAC	5100
AACCATGATC	ATGGCATCTT	TCTGCCAAAT	AGATATCACA	AAACTACAGC	ACGTAGGTGT	5160
ATTGGTCAAC	AGTCCCAAAG	GGATGAAGAT	ATCACAAAAC	TTTGAAACAA	GATATCTAAT	5220
TTTGAGCCTC	ATACCAAAAA	TAGAAGACTC	TAACTCTTGT	GGTGACCAAC	AGATCAAGCA	5280
ATACAAGAAG	TTATTGGATA	GACTGATCAT	CCCTTTATAT	GATGGATTAA	GATTACAGAA	5340
AGATGTGATA	GTAACCAATC	AAGAATCCAA	TGAAAACACT	GATCCCAGAA	CAAAACGATT	5400
CTTTGGAGGG	GTAATTGGAA	CCATTGCTCT	GGGAGTAGCA	ACCTCAGCAC	AAATTACAGC	5460
GGCAGTTGCT	CTGGTTGAAG	CCAAGCAGGC	AAGATCAGAC	ATCGAAAAAC	TCAAAGAAGC	5520
AATTAGGGAC	ACAAATAAAG	CAGTGCAGTC	AGTTCAGAGC	TCCATAGGAA	ATTTAATAGT	5580
AGCAATTAAA	TCAGTCCAGG	ATTATGTTAA	CAAAGAAATC	GTGCCATCGA	TTGCGAGGCT	5640
AGGTTGTGAA	GCAGCAGGAC	TTCAATTAGG	AATTGCATTA	ACACAGCATT	ACTCAGAATT	5700
AACAAACATA	TTTGGTGATA	ACATAGGATC	GTTACAAGAA	AAAGGAATAA	AATTACAAGG	5760
TATAGCATCA	TTATACCGCA	CAAATATCAC	AGAAATATTC	ACAACATCAA	CAGTTGATAA	5820
ATATGATATC	TATGATCTGT	TATTTACAGA	ATCAATAAAG	GTGAGAGTTA	TAGATGTTGA	5880
CTTGAATGAT	TACTCAATCA	CCCTCCAAGT	CAGACTCCCT	TTATTAACTA	GGCTGCTGAA	5940
CACTCAGATC	TACAAAGTAG	ATTCCATATC	ATATAACATC	CAAAACAGAG	AATGGTATAT	6000
CCCTCTTCCC	AGCCATATCA	TGACGAAAGG	GGCATTTCTA	GGTGGAGCAG	acgtcaaaga	6060

- 259 -

ATGTATAGAA	GCATTCAGCA	GCTATATATG	CCCTTCTGAT	CCAGGATTTG	TATTAAACCA	6120
TGAAATAGAG	AGCTGCTTAT	CAGGAAACAT	ATCCCAATGT	CCAAGAACAA	CGGTCACATC	6180
AGACATTGTT	CCAAGATATG	CATTTGTCAA	TGGAGGAGTG	GTTGCAAACT	GTATAACAAC	6240
CACCTGTACA	TGCAACGGAA	TTGGTAATAG	AATCAATCAA	CCACCTGATC	AAGGAGTAAA	6300
AATTATAACA	CATAAAGAAT	GTAGTACAGT	AGGTATCAAC	GGAATGCTGT	TCAATACAAA	6360
TAAAGAAGGA	ACTCTTGCAT	TCTATACACC	AAATGATATA	ACACTAAACA	ATTCTGTTAC	6420
ACTTGATCCA	ATTGACATAT	CAATCGAGCT	CAACAAGGCC	AAATCAGATC	TAGAAGAATC	6480
AAAAGAATGG	ATAAGAAGGT	CAAATCAAAA	ACTAGATTCT	ATTGGAAATT	GGCATCAATC	6540
TAGCACTACA	ATCATAATTA	TTTTGATAAT	GATCATTATA	TTGTTTATAA	TTAATATAAC	6600
GATAATTACA	ATTGCAATTA	AGTATTACAG	AATTCAAAAG	AGAAATCGAG	TGGATCAAAA	6660
TGACAAGCCA	TATGTACTAA	CAAACAAATA	ACATATCTAC	AGATCATTAG	ATATTAAAAT	6720
TATAAAAAAC	TTAGGAGTAA	AGTTACGCAA	TCCAACTCTA	CTCATATAAT	TGAGGAAGGA	6780
CCCAATAGAC	AAATCCAAAT	TCGAGATGGA	ATACTGGAAG	CATACCAATC	ACGGAAAGGA	6840
TGCTGGCAAT	GAGCTGGAGA	CGTCTATGGC	TACTCATGGC	AACAAGCTCA	CTAATAAGAT	6900
AATATACATA	TTATGGACAA	TAATCCTGGT	GTTATTATCA	ATAGTCTTCA	TCATAGTGCT	6960
AATTAATTCC	ATCAAAAGTG	AAAAGGCCCA	CGAATCATTG	CTGCAAGACA	TAAATAATGA	7020
					ATCTAATACA	7080
GTCAGGAGTG	AATACAAGGC	TTCTTACAAT	TCAGAGTCAT	GTCCAGAATT	ACATACCAAT	7140
ATCATTGACA	CAACAGATGT	CAGATCTTAG	GAAATTCATT	AGTGAAATTA	CAATTAGAAA	7200
					AACCTTTAAA	7260
			TCTTCCATCT			7320
					GCTGTGTTAG	
					TAATTACTCG	
					TAACTGTAAA	
					TAAATGACAA	
TAGGAAGTCA	TGTTCTCTAG	CACTCCTAAA	TACAGATGTA	TATCAACTGT	GTTCAACTCC	7620

- 260 -

CAAAGTTGAT	GAAAGATCAG	ATTATGCATC	ATCAGGCATA	GAAGATATTG	TACTTGATAT	7680
TGTCAATTAT	GATGGTTCAA	TCTCAACAAC	AAGATTTAAG	AATAATAACA	TAAGCTTTGA	7740
TCAACCATAT	GCTGCACTAT	ACCCATCTGT	TGGACCAGGG	ATATACTACA	AAGGCAAAAT	7800
AATATTTCTC	GGGTATGGAG	GTCTTGAACA	TCCAATAAAT	GAGAATGTAA	TCTGCAACAC	7860
AACTGGGTGC	CCCGGGAAAA	CACAGAGAGA	CTGTAATCAA	GCGTCTCATA	GTCCATGGTT	7920
TTCAGATAGG	AGGATGGTCA	ACTCCATCAT	TGTTGCTGAC	AAAGGCTTAA	ACTCAATTCC	7980
AAAATTGAAA	GTATGGACGA	TATCTATGCG	ACAAAATTAC	TGGGGGTCAG	AAGGAAGGTT	8040
ACTTCTACTA	GGTAACAAGA	TCTATATATA	TACAAGATCT	ACAAGTTGGC	ATAGCAAGTT	8100
ACAATTAGGA	ATAATTGATA	TTACTGATTA	CAGTGATATA	AGGATAAAAT	GGACATGGCA	8160
TAATGTGCTA	TCAAGACCAG	GAAACAATGA	ATGTCCATGG	GGACATTCAT	GTCCAGATGG	8220
ATGTATAACA	GGAGTATATA	CTGATGCATA	TCCACTCAAT	CCCACAGGGA	GCATTGTGTC	8280
ATCTGTCATA	TTAGACTCAC	AAAAATCGAG	AGTGAACCCA	GTCATAACTT	ACTCAACAGC	8340
AACCGAAAGA	GTAAACGAGC	TGGCCATCCT	AAACAGAACA	CTCTCAGCTG	GATATACAAC	8400
AACAAGCTGC	ATTACACACT	ATAACAAAGG	ATATTGTTTT	CATATAGTAG	AAATAAATCA	8460
TAAAAGCTTA	AACACATTTC	AACCCATGTT	GTTCAAAACA	GAGATTCCAA	AAAGCTGCAG	8520
TTAATCATAA	TTAACCATAA	TATGCATCAA	TCTATCTATA	ATACAAGTAT	ATGATAAGTA	8580
ATCAGCAATC	AGACAATAGA	CAAAAGGGAA	ATATAAAAA	CTTAGGAGCA	AAGCGTGCTC	8640
gggaaatgga	CACTGAATCT	AACAATGGCA	CTGTATCTGA	CATACTCTAT	CCTGAGTGTC	8700
ACCTTAACTC	TCCTATCGTT	AAAGGTAAAA	TAGCACAATT	ACACACTATT	ATGAGTCTAC	8760
CTCAGCCTTA	TGATATGGAT	GACGACTCAA	TACTAGTTAT	CACTAGACAG	AAAATAAAAC	8820
TTAATAAATT	ggataaaaga	CAACGATCTA	TTAGAAGATT	ATAATTAAAA	TTAACTGAAA	8880
ADTAADTDAA	CTTAGGAAAA	TACACATTTA	TCAGATATCC	AGAAATGTCA	AAAGAAATGT	8940
TCAAATTATA	TATACCTGGT	ATTAACAGTA	AAGTGACTGA	ATTATTACTT	AAAGCAGATA	9000
GAACATATAG	TCAAATGACT	GATGGATTAA	GAGATCTATG	GATTAATGTG	CTATCAAAAT	9060
TAGCCTCAAA	AAATGATGGA	AGCAATTATG	ATCTTAATGA	AGAAATTAAT	AATATATCGA	9120
AAGTTCACAC	AACCTATAAA	TCAGATAAAT	GGTATAATCC	ATTCAAAACA	TGGTTTACTA	9180

- 261 -

TCAAGTATGA	TATGAGAAGA	TTACAAAAAG	CTCGAAATGA	GATCACTTTT	AATGTTGGGA	9240
AGGATTATAA	CTTGTTAGAA	GACCAGAAGA	ATTTCTTATT	GATACATCCA	GAATTGGTTT	9300
TGATATTAGA	TARACARARC	TACAATGGTT	ATCTAATTAC	TCCTGAATTA	GTATTGATGT	9360
ATTGTGACGT	AGTCGAAGGC	CGATGGAATA	TAAGTGCATG	TGCTAAGTTA	GATCCAAAAT	9420
TACAATCTAT	GTATCAGAAA	GGTAATAACC	TGTGGGAAGT	GATAGATAAA	TTGTTTCCAA	9480
TTATGGGAGA	AAAGACATTT	GATGTGATAT	CGTTATTAGA	ACCACTTGCA	TTATCCTTAA	9540
TTCAAACTCA	TGATCCTGTT	AAACAACTAA	GAGGAGCTTT	TTTAAATCAT	GTGTTATCCG	9600
agatggaatt	AATATTTGAA	TCTAGAGAAT	CGATTAAGGA	ATTTCTGAGT	GTAGATTACA	9660
TTGATAAAAT	TTTAGATATA	TTTAATAAGT	CTACAATAGA	TGAAATAGCA	GAGATTTTCT	9720
CTTTTTTAG	AACATTTGGG	CATCCTCCAT	TAGAAGCTAG	TATTGCAGCA	GAAAAGGTTA	9780
GAAAATATAT	GTATATTGGA	AAACAATTAA	AATTTGACAC	TATTAATAAA	TGTCATGCTA	9840
TCTTCTGTAC	AATAATAATT	AACGGATATA	GAGAGAGGCA	TGGTGGACAG	TGGCCTCCTG	9900
TGACATTACC	TGATCATGCA	CACGAATTCA	TCATAAATGC	TTACGGTTCA	AACTCTGCGA	9960
TATCATATGA	GAATGCTGTT	GATTATTACC	AGAGCTTTAT	AGGAATAAAA	TTCAATAAAT	10020
TCATAGAGCC	TCAGTTAGAT	GAGGATTTGA	CAATTTATAT	GAAAGATAAA	GCATTATCTC	10080
CAAAAAAATC	AAATTGGGAC	ACAGTTTATC	CTGCATCTAA	TTTACTGTAC	CGTACTAACG	10140
CATCCAACGA	ATCACGAAGA	TTAGTTGAAG	TATTTATAGC	agatagtaaa	TTTGATCCTC	10200
ATCAGATATT	GGATTATGTA	GAATCTGGGG	ACTGGTTAGA	TGATCCAGAA	TTTAATATTT	10260
CTTATAGTCT	TAAAGAAAAA	GAGATCAAAC	AGGAAGGTAG	ACTCTTTGCA	AAAATGACAT	10320
ACAAAATGAG	AGCTACACAA	GTTTTATCAG	AGACACTACT	TGCAAATAAC	ATAGGAAAAT	10380
TCTTTCAAGA	AAATGGGATG	GTGAAGGGAG	AGATTGAATT	ACTTAAGAGA	TTAACAACCA	10440
TATCAATATC	AGGAGTTCCA	CGGTATAATG	AAGTGTACAA	TAATTCTAAA	AGCCATACAG	10500
ATGACCTTAA	AACCTACAAT	AAAATAAGTA	ATCTTAATTT	GTCTTCTAAT	CAGAAATCAA	10560
AGAAATTTGA	ATTCAAGTCA	ACGGATATCT	ACAATGATGG	ATACGAGACT	GTGAGCTGTT	10620
TCCTAACAAC	AGATCTCAAA	AAATACTGTC	TTAATTGGAG	ATATGAATCA	ACAGCTCTAT	10680
TTGGAGAAAC	TTGCAACCAA	ATATTTGGAT	TAAATAAATT	GTTTAATTGG	TTACACCCTC	10740

GTCTTGAAGG	AAGTACAATC	TATGTAGGT	ATCCTTACTO	F TCCTCCATC	A GATAAAGAAC	10800
ATATATCATT	AGAGGATCAC	CCTGATTCTG	GTTTTTACGT	TCATAACCC	A AGAGGGGGTA	10860
TAGAAGGATT	TTGTCAAAAA	TTATGGACAC	: TCATATCTAT	AAGTGCAAT	CATCTAGCAG	10920
CTGTTAGAAT	AGGCGTGAGG	GTGACTGCAA	TGGTTCAAGG	AGACAATCAI	GCTATAGCTG	10980
TAACCACAAG	AGTACCCAAC	AATTATGACT	' ACAGAGTTAA	GAAGGAGATI	GTTTATAAAG	11040
atgtagtgag	ATTTTTTGAT	TCATTAAGAG	AAGTGATGGA	TGATCTAGG1	CATGAACTTA	11100
AATTAAATGA	AACGATTATA	AGTAGCAAGA	TGTTCATATA	TAGCAAAAGA	ATCTATTATG	11160
ATGGGAGAAT	TCTTCCTCAA	GCTCTAAAAG	CATTATCTAG	ATGTGTCTTC	: TGGTCAGAGA	11220
CAGTAATAGA	CGAAACAAGA	TCAGCATCTT	CAAATTTGGC	AACATCATTT	GCAAAAGCAA	11280
TTGAGAATGG	TTATTCACCT	GTTCTAGGAT	ATGCATGCTC	AATTTTTAAG	AACATTCAAC	11340
AACTATATAT	TGCCCTTGGG	ATGAATATCA	ATCCAACTAT	AACACAGAAT	ATCAGAGATC	11400
AGTATTTAG	GAATCCAAAT	TGGATGCAAT	ATGCCTCTTT	AATACCTGCT	AGTGTTGGGG	11460
GATTCAATCA	CATGGCCATG	TCAAGATGTT	TTGTAAGGAA	TATTGGTGAT	CCATCAGTTG	11520
CCGCATTGGC	TGATATTAAA	AGATTTATTA	AGGCGAATCT	ATTAGACCGA	AGTGTTCTTT	11580
ATAGGATTAT	GAATCAAGAA	CCAGGTGAGT	CATCTTTTTT	TGACTGGGCT	TCAGATCCAT	11640
ATTCATGCAA	TTTACCACAA	TCTCAAAATA	TAACCACCAT	GATAAAAAAT	ATAACAGCAA	11700
GGAATGTATT	ACAAGATTCA	CCAAATCCAT	TATTATCTGG	ATTATTCACA	AATACAATGA	11760
TAGAAGAAGA	TGAAGAATTA	GCTGAGTTCC	TGATGGACAG	GAAGGTAATT	CTCCCTAGAG	11820
TTGCACATGA	TATTCTAGAT	AATTCTCTCA	CAGGAATTAG	AAATGCCATA	GCTGGAATGT	11880
TAGATACGAC	AAAATCACTA	ATTCGGGTTG	GCATAAATAG	AGGAGGACTG	ACATATAGTT	11940
TGTTGAGGAA	AATCAGTAAT	TACGATCTAG	TACAATATGA	AACACTAAGT	AGGACTTTGC	12000
GACTAATTGT	AAGTGATAAA	ATCAAGTATG	AAGATATGTG	TTCGGTAGAC	CTTGCCATAG	12060
CATTGCGACA	Aaagatgtgg	ATTCATTTAT	CAGGAGGAAG	GATGATAAGT	GGACTTGAAA	12120
CGCCTGACCC	ATTAGAATTA	CTATCTGGGG	TAGTAATAAC	AGGATCAGAA	CATTGTAAAA	12180
TATGTTATTC	TTCAGATGGC	ACAAACCCAT	ATACTTGGAT	GTATTTACCC	GGTAATATCA	12240
AAATAGGATC	AGCAGAAACA	GGTATATCGT	CATTAAGAGT	TCCTTATTTT	GGATCAGTCA	12300

- 263 -

CTGATGAAAG A	ATCTGAAGCA	CAATTAGGAT	ATATCAAGAA	TCTTAGTAAA	CCTGCAAAAG	12360
CCGCAATAAG A	ATAGCAATG	ATATATACAT	GGGCATTTGG	TAATGATGAG	ATATCTTGGA	12420
TGGAAGCCTC A	CAGATAGCA	CAAACACGTG	CAAATTTTAC	ACTAGATAGT	CTCAAAATTT	12480
TAACACCGGT A	GCTACATCA	ACAAATTTAT	CACACAGATT	AAAGGATACT	GCAACTCAGA	12540
TGAAATTCTC C	agtacatca	TTGATCAGAG	TCAGCAGATT	TATAACAATG	TCCAATGATA	12600
ACATGTCTAT C	AAAGAAGCT	AATGAAACCA	AAGATACTAA	TCTTATTTAT	CAACAAATAA	12660
TGTTAACAGG A	TTAAGTGTT	TTCGAATATT	TATTTAGATT	AAAAGAAACC	ACAGGACACA	12720
ACCCTATAGT T	ATGCATCTG	CACATAGAAG	ATGAGTGTTG	TATTAAAGAA	AGTTTTAATG	12780
ATGAACATAT T	AATCCAGAG	TCTACATTAG	AATTAATTCG	ATATCCTGAA	AGTAATGAAT	12840
TTATTTATGA T	AAAGACCCA	CTCAAAGATG	TGGACTTATC	AAAACTTATG	GTTATTAAAG	12900
ACCATTCTTA C	ACAATTGAT	ATGAATTATT	GGGATGATAC	TGACATCATA	CATGCAATTT	12960
CAATATGTAC T	GCAATTACA	ATAGCAGATA	CTATGTCACA	ATTAGATCGA	GATAATTTAA	13020
AAGAGATAAT A	GTTATTGCA	AATGATGATG	ATATTAATAG	CTTAATCACT	GAATTTTTGA	13080
CTCTTGACAT A	CTTGTATTT	CTCAAGACAT	TTGGTGGATT	ATTAGTAAAT	CAATTTGCAT	13140
ACACTCTTTA T	AGTCTAAAA	Atagaaggta	GGGATCTCAT	TTGGGATTAT	ATAATGAGAA	13200
CACTGAGAGA T	ACTTCCCAT	TCAATATTAA	AAGTATTATC	TAATGCATTA	TCTCATCCTA	13260
AAGTATTCAA G	AGGTTCTGG	GATTGTGGAG	TTTTAAACCC	TATTTATGGT	CCTAATATTG	13320
CTAGTCAAGA C	CAGATAAAA	CTTGCCCTAT	CTATATGTGA	ATATTCACTA	GATCTATTTA	13380
TGAGAGAATG G	TTGAATGGT	GTATCACTTG	AAATATACAT	TTGTGACAGC	GATATGGAAG	13440
TTGCAAATGA T.	AGGAAACAA	GCCTTTATTT	CTAGACACCT	TTCATTTGTT	TGTTGTTTAG	13500
CAGAAATTGC A	TCTTTCGGA	CCTAACCTGT	TAAACTTAAC	ATACTTGGAG	AGACTTGATC	13560
TATTGAAACA A	TATCTTGAA	TTAAATATTA	AAGAAGACCC	TACTCTTAAA	TATGTACAAA	13620
TATCTGGATT A	AAATTAAT	TCGTTCCCAT	CAACTGTAAC	ATACGTAAGA	AAGACTGCAA	13680
TCAAATATCT A	AGGATTCGC (	GGTATTAGTC	CACCTGAGGT	AATTGATGAT	TGGGATCCGG	13740
TAGAAGATGA A	AATATGCTG (	GATAACATTG	TCAAAACTAT	AAATGATAAC	TGTAATAAAG	13800
ATAATAAAGG G	AATAAAATT :	AACAATTTCT	GGGGACTAGC	ACTTAAGAAC	TATCAAGTCC	13860

- 264 -

	•					
TTAAAATCAG	ATCTATAACA	AGTGATTCT	atgataatga	TAGACTAGAT	GCTAATACAA	13920
GTGGTTTGAC	ACTTCCTCAR	GGAGGGAATT	ATCTATCGCA	TCAATTGAGA	TTATTCGGAA	13980
TCAACAGCAC	TAGTTGTCTG	AAAGCTCTTG	AGTTATCACA	AATTTTAATG	AAGGAAGTCA	14040
ATAAAGACAA	GGACAGGCTC	TTCCTGGGAG	AAGGAGCAGG	AGCTATGCTA	GCATGTTATG	14100
ATGCCACATT	AGGACCTGCA	GTTAATTATT	' ATAATTCAGG	TTTGAATATA	ACAGATGTAA	14160
TTGGTCAACG	AGAATTGAAA	ATATTTCCTT	CAGAGGTATC	ATTAGTAGGT	AAAAAATTAG	14220
GAAATGTGAC	ACAGATTCTT	AACAGGGTAA	AAGTACTGTT	CAATGGGAAT	CCTAATTCAA	14280
CATGGATAGG	AAATATGGAA	TGTGAGAGCT	TAATATGGAG	TGAATTAAAT	GATAAGTCCA	14340
TTGGATTAGT	ACATTGTGAT	ATGGAAGGAG	CTATCGGTAA	ATCAGAAGAA	ACTGTTCTAC	14400
ATGAACATTA	TAGTGTTATA	AGAATTACAT	ACTTGATTGG	GGATGATGAT	GTTGTTTTAG	14460
TTTCCAAAAT	TATACCTACA	ATCACTCCGA	ATTGGTCTAG	AATACTTTAT	CTATATAAAT	14520
TATATTGGAA	AGATGTAAGT	ATAATATCAC	TCAAAACTTC	TAATCCTGCA	TCAACAGAAT	14580
TATATCTAAT	TTCGAAAGAT	GCATATTGTA	CTATAATGGA	ACCTAGTGAA	ATTGTTTTAT	14640
CAAAACTTAA	AAGATTGTCA	CTCTTGGAAG	AAAATAATCT	ATTAAAATGG	ATCATTTAT	14700
CAAAGAAGAG	Gaataatgaa	TGGTTACATC	ATGAAATCAA	AGAAGGAGAA	AGAGATTATG	14760
GAATCATGAG	ACCATATCAT	ATGGCACTAC	AAATCTTTGG	ATTTCAAATC	AATTTAAATC	14820
ATCTGGCGAA	ATTTTTAADA	TCAACCCCAG	ATCTGACTAA	TATCAACAAT	ATAATCCAAA	14880
GTTTTCAGCG	AACAATAAAG	GATGTTTTAT	TTGAATGGAT	TAATATAACT	CATGATGATA	14940
AGAGACATAA	ATTAGGCGGA	AGATATAACA	TATTCCCACT	GAAAAATAAG	GGAAAGTTAA	15000
GACTGCTATC	GAGAAGACTA	GTATTAAGTT	GGATTTCATT	ATCATTATCG	ACTCGATTAC	15060
TTACAGGTCG	CTTTCCTGAT	GAAAAATTTG	AACATAGAGC	ACAGACTGGA	TATGTATCAT	15120
TAGCTGATAC	TGATTTAGAA	TCATTAAAGT	TATTGTCGAA	AAACATCATT	AAGAATTACA	15180
GAGAGTGTAT	AGGATCAATA	TCATATTGGT	TTCTAACCAA	AGAAGTTAAA	ATACTTATGA	15240
AATTGATTGG	TGGTGCTAAA	TTATTAGGAA	TTCCCAGACA	ATATAAAGAA	CCCGAAGACC	15300
AGTTATTAGA	AAACTACAAT	CAACATGATG	AATTTGATAT	CGATTAAAAC	ATAAATACAA	15360
TGAAGATATA	TCCTAACCTT	TATCTTTAAG	CCTAGGAATA	GACAAAAAGT	AAGAAAAACA	15420

- 265 -

## TGTAATATAT ATATACCAAA CAGAGTTCTT CTCTTGTTTG GT

15462

- (2) INFORMATION FOR SEQ ID NO:20:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2233 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
  - Met Asp Thr Glu Ser Asn Asn Gly Thr Val Ser Asp Ile Leu Tyr Pro 1 5 10 15
  - Glu Cys His Leu Asn Ser Pro Ile Val Lys Gly Lys Ile Ala Gln Leu 20 25 30
  - His Thr Ile Met Ser Leu Pro Gln Pro Tyr Asp Met Asp Asp Asp Ser 35 40 45
  - Ile Leu Val Ile Thr Arg Gln Lys Ile Lys Leu Asn Lys Leu Asp Lys 50 55 60
  - Arg Gln Arg Ser Ile Arg Arg Leu Lys Leu Ile Leu Thr Glu Lys Val 65 70 75 80
  - Asn Asp Leu Gly Lys Tyr Thr Phe Ile Arg Tyr Pro Glu Met Ser Lys 85 90 95
  - Glu Met Phe Lys Leu Tyr Ile Pro Gly Ile Asn Ser Lys Val Thr Glu 100 105 110
  - Leu Leu Lys Ala Asp Arg Thr Tyr Ser Gln Met Thr Asp Gly Leu 115 120 125
  - Arg Asp Leu Trp Ile Asn Val Leu Ser Lys Leu Ala Ser Lys Asn Asp 130 135 140
  - Gly Ser Asn Tyr Asp Leu Asn Glu Glu Ile Asn Asn Ile Ser Lys Val 145 150 155 160
  - His Thr Thr Tyr Lys Ser Asp Lys Trp Tyr Asn Pro Phe Lys Thr Trp

    165 170 175
  - Phe Thr Ile Lys Tyr Asp Met Arg Arg Leu Gln Lys Ala Arg Asn Glu

- 266 -

			180					185					190		
Ile	Thr	Phe 195	Asn	Val	Gly	Lys	<b>Asp</b> 200	Tyr	Asn	Leu	Leu	Glu 205	Двр	Gln	Lye
Asn	Phe 210	Leu	Leu	Ile	His	Pro 215	Glu	Leu	Val	Leu	Ile 220	Leu	Asp	Lys	Glr
Asn 225	Tyr	Asn	Gly	Tyr	Leu 230	Ile	Thr	Pro	Glu	<b>Le</b> u 235	Val	Leu	Met	Tyr	Cys 240
Asp	Val	Val	Glu	Gly 245	Arg	Trp	Asn	Ile	Ser 250	Ala	Сув	Ala	Lys	Leu 255	Asp
Pro	Lys	Leu	Gln 260	Ser	Met	Tyr	Gln	<b>Lys</b> 265	Gly	Asn	Asn	Leu	Trp 270	Glu	Val
Ile	Asp	Lys 275	Leu	Phe	Pro	Ile	Met 280	Gly	Glu	Lys	Thr	Phe 285	Asp	Val	Ile
Ser	Leu 290	Leu	Glu	Pro	Leu	Ala 295	Leu	Ser	Leu	Ile	Gln 300	Thr	His	Asp	Pro
Val 305	Lys	Gln	Leu	Arg	Gly 310	Ala	Phe	Leu	Asn	His 315	Val	Leu	Ser	Glu	Met 320
Glu	Leu	Ile	Phe	Glu 325	Ser	Arg	Glu	Ser	11e 330	Lys	Glu	Phe	Leu	Ser 335	Val
Asp	Tyr	Ile	<b>Авр</b> 340	Lys	Ile	Leu	Авр	Ile 345	Phe	Asn	Lys	Ser	Thr 350	Ile	Asp
Glu	Ile	Ala 355	Glu	Ile	Phe	Ser	Phe 360	Phe	Arg	Thr	Phe	Gly 365	His	Pro	Pro
Leu	Glu 370	Ala	Ser	Ile	Ala	Ala 375	Glu	Lys	Val	Arg	Lys 380	Tyr	Met	Tyr	Ile
Gly 385	Lys	Gln	Leu	Lys	Phe 390	Asp	Thr	Ile	Asn	Lys 395	Сув	His	Ala	Ile	Phe 400
Сув	Thr	Ile	Ile	Ile 405	Asn	Gly	Tyr	Arg	Glu 410	Arg	His	Gly	Gly	Gln 415	Trp
Pro	Pro	Val	Thr 420	Leu	Pro	Asp	His	Ala 425	His	Glu	Phe	Ile	Ile 430	Asn	Ala
Tyr	Gly	Ser 435	Asn	Ser	Ala	Ile	Ser 440	Tyr	Glu	Asn	Ala	Val 445	Asp	Tyr	Tyr
Gln	Ser 450	Phe	Ile	Gly	Ile	Lys 455	Phe	Asn	Lys	Phe	Ile 460	Glu	Pro	Gln	Leu

- 267 -

<b>Asp</b>		Asp	Leu	Thr	11e 470		Met	Lys	Asr	475		. Let	ı Ser	Pro	480
Lys	Ser	Asn	Trp	Asp 485	Thr	Val	Tyr	Pro	Ala 490		. Asr	Leu	1 Leu	1 Tyr 495	
Thr	Asn	Ala	Ser 500		Glu	Ser	Arg	Arg 505		ı Val	. Glu	Val	Phe 510		Ala
Asp	Ser	Lys 515		Авр	Pro	His	Gln 520		Lev	. Asp	Туг	Val		Ser	Gly
Asp	Trp 530	Leu	Авр	Авр	Pro	Glu 535		Asn	Ile	Ser	Tyr 540		Leu	Lys	Glu
<b>Lys</b> 545	Glu	Ile	Lys	Gln	Glu 550	Gly	Arg	Leu	Phe	Ala 555		Met	Thr	Tyr	Lys 560
Met	Arg	Ala	Thr	Gln 565	Val	Leu	Ser	Glu	Thr 570		Leu	Ala	Asn	Asn 575	
Gly	Lув	Phe	Phe 580	Gln	Glu	Asn	Gly	Met 585		Lys	Gly	Glu	11e 590	Glu	Leu
Leu	Lys	<b>Arg</b> 595	Leu	Thr	Thr	Ile	Ser 600	Ile	Ser	Gly	Val	Pro 605	_	Tyr	Asn
Glu	Val 610	Tyr	Asn	Asn	Ser	Lув 615	Ser	His	Thr	qaA	Asp 620	Leu	Lys	Thr	Tyr
Asn 625	Lys	Ile	Ser	Asn	Leu 630	Asn	Leu	Ser	Ser	Asn 635	Gln	Lys	Ser	Lys	Lys 640
Phe	<b>Gl</b> u	Phe	Lys	Ser 645	Thr	<b>As</b> p	Ile	Tyr	Asn 650	Asp	Gly	Tyr	Glu	Thr 655	Val
Ser	Сув	Phe	660 Leu	Thr	Thr	Asp	Leu	<b>L</b> ув 665	Lys	Tyr	Сув	Leu	<b>Asn</b> 670	Trp	Arg
Tyr	Glu	Ser 675	Thr	Ala	Leu	Phe	Gly 680	Glu	Thr	Сув	Asn	Gln 685	Ile	Phe	Gly
Leu	Asn 690	Lys	Leu	Phe	Asn	Trp 695	Leu	His	Pro	Arg	Leu 700	Glu	Gly	Ser	Thr
11e 705	Tyr	Val	Gly	Asp	Pro 710	Tyr	Сув	Pro	Pro	Ser 715	Asp	Lys	Glu	His	Ile 720
Ser	Leu	Glu	Asp	His 725	Pro	Asp	Ser	Gly	Phe 730	Tyr	Val	His	Asn	Pro 735	Arg

- 268 -

Gly	Gly	Ile	Glu 740	Gly	Phe	Сув	Gln	Lys 745		Trp	Thr	Leu	750		Ile
Ser	Ala	Ile 755	His	Leu	Ala	Ala	Val 760		Ile	Gly	Val	Arg 765		Thr	Ala
Met	Val 770	Gln	Gly	Asp	Asn	Gln 775		Ile	Ala	Val	Thr 780		Arg	Val	Pro
Asn 785	Asn	Tyr	Asp	Tyr	Arg 790	Val	Lys	Lys	Glu	Ile 795		Tyr	Lys	Asp	Val 800
Val	Arg	Phe	Phe	<b>А</b> вр 805	Ser	Leu	Arg	Glu	Val 810		Asp	Asp	Leu	Gly 815	His
Glu	Leu	Lys	Leu 820	Asn	Glu	Thr	Ile	Ile 825		Ser	Lys	Met	Phe 830	Ile	Tyr
Ser	Lys	Arg 835	Ile	Tyr	Tyr	Asp	Gly 840	Arg	Ile	Leu	Pro	Gln 845	Ala	Leu	Lys
Ala	<b>Leu</b> 850	Ser	Arg	Сув	Val	Phe 855	Trp	Ser	Glu	Thr	Val 860	Ile	Авр	Glu	Thr
Arg 865	Ser	Ala	Ser	Ser	Asn 870	Leu	Ala	Thr	Ser	Phe 875	Ala	Lys	Ala	Ile	Glu 880
Asn	Gly	Tyr	Ser	Pro 885	Val	Leu	Gly	Tyr	Ala 890	Сув	Ser	Ile	Phe	Lys 895	Asn
Ile	Gln	Gln	Leu 900	туг	Ile	Ala	Leu	Gly 905	Met	Asn	Ile	Asn	Pro 910	Thr	Ile
Thr	Gln	Asn 915	Ile	Arg	Asp	Gln	Туг 920	Phe	Arg	Asn	Pro	Asn 925	Trp	Met	Gln
Tyr	Ala 930	Ser	Leu	Ile	Pro	Ala 935	Ser	Val	Gly	Gly	Phe 940	Asn	His	Met	Ala
Met 945	Ser	Arg	Сув	Phe	Val 950	Arg	Asn	Ile	Gly	Asp 955	Pro	Ser	Val	Ala	Ala 960
Leu	Ala	Авр	Ile	Lys 965	Arg	Phe	Ile	Lys	Ala 970	Asn	Leu	Leu	Asp	Arg 975	Ser
Val	Leu	Tyr	Arg 980	Ile	Met	Asn	Gln	Glu 985	Pro	Gly	Glu	Ser	Ser 990	Phe	Phe
		995	Ser				1000	)				1005	<b>5</b>		
Ile	Thr	Thr	Met	Ile	Lys	Asn	Ile	Thr	Ala	Arg	Asn	Val	Leu	Gln	Asp

- 269 -

	1010	)				1015	5 '				102	0			
Ser 1025		Asn	Pro	Leu	Leu 103(		Gly	Leu	Phe	Thr 103		Thr	Met	Ile	Glu 1040
Glu	Asp	Glu	Glu	Leu 1045		Glu	Phe	Leu	Met 105	-	Arg	Lys	Val	Ile 105	Leu 5
Pro	Arg	Val	Ala 1060		Asp	Ile	Leu	Asp 1069		Ser	Leu	Thr	Gly 107		λrg
Asn	Ala	Ile 1075		Gly	Met	Leu	<b>Asp</b>		Thr	Lys	Ser	Leu 108		Arg	Val
	Ile 1090	Asn	Arg	Gly	Gly	Leu 1095		Tyr	Ser	Leu	Leu 110	_	Lys	Ile	Ser
Asn 1105		Asp	Leu	Val	Gln 1110		Glu	Thr	Leu	Ser 1115		Thr	Leu	Arg	Leu 1120
Ile	Val	Ser	Asp	Lys 1125		Lys	Tyr	Glu	Авр 1130		Сув	Ser	Val	Asp 1135	Leu
Ala	Ile	Ala	Leu 1140		Gln	Lys	Met	Trp 1145		His	Leu	Ser	Gly 1150	_	Arg
Met	Ile	Ser 1155		Leu	Glu	Thr	Pro 1160		Pro	Leu	Glu	Leu 1165		Ser	Gly
	Val 1170	Ile	Thr	Gly	Ser	Glu 1175		Сув	Lys	Ile	Сув 1180	-	Ser	Ser	Asp
3ly 1185		Asn	Pro		Thr 1190		Met	Tyr	Leu	Pro 1195	_	Asn	Ile	Lys	Ile 1200
31y	Ser	Ala		Thr 1205		Ile	Ser	Ser	Leu 1210		Val	Pro	Tyr	Phe 1215	
Ser '	Val	Thr	Авр 1220		Arg	Ser		Ala 1225		Leu	Gly	Tyr	Ile 1230	_	Asn
Leu .		Lys 1235		Ala	Lys		Ala 1240		Arg	Ile	Ala	Met 1245		Tyr	Thr
	Ala 1250	Phe	Gly	an .		Glu 1255		Ser	Trp		Glu 1260		Ser	Gln	Ile
l265	Gln	Thr .	Arg		Asn 1270	Phe '	Thr	Leu		Ser 1275		Lys	Ile		Thr 1280
Pro '	Val .	Ala '		Ser ' 1285	Thr .	Asn :	Leu		His 1290		Leu	Lys		Thr 1295	

- Thr Gln Met Lys Phe Ser Ser Thr Ser Leu Ile Arg Val Ser Arg Phe 1300 · 1305 1310
- Ile Thr Met Ser Asn Asp Asn Met Ser Ile Lys Glu Ala Asn Glu Thr 1315 1320 1325
- Lys Asp Thr Asn Leu Ile Tyr Gln Gln Ile Met Leu Thr Gly Leu Ser 1330 1335 1340
- Val Phe Glu Tyr Leu Phe Arg Leu Lys Glu Thr Thr Gly His Asn Pro 1345 1350 1355 1360
- Ile Val Met His Leu His Ile Glu Asp Glu Cys Cys Ile Lys Glu Ser 1365 1370 1375
- Phe Asn Asp Glu His Ile Asn Pro Glu Ser Thr Leu Glu Leu Ile Arg 1380 1385 1390
- Tyr Pro Glu Ser Asn Glu Phe Ile Tyr Asp Lys Asp Pro Leu Lys Asp 1395 1400 1405
- Val Asp Leu Ser Lys Leu Met Val Ile Lys Asp His Ser Tyr Thr Ile 1410 1415 1420
- Asp Met Asn Tyr Trp Asp Asp Thr Asp Ile Ile His Ala Ile Ser Ile 1425 1430 1435 1446
- Cys Thr Ala Ile Thr Ile Ala Asp Thr Met Ser Gln Leu Asp Arg Asp 1445 1450 1455
- Asn Leu Lys Glu Ile Ile Val Ile Ala Asn Asp Asp Asp Ile Asn Ser 1460 1465 1470
- Leu Ile Thr Glu Phe Leu Thr Leu Asp Ile Leu Val Phe Leu Lys Thr 1475 1480 1485
- Phe Gly Gly Leu Leu Val Asn Gln Phe Ala Tyr Thr Leu Tyr Ser Leu 1490 1495 1500
- Lys Ile Glu Gly Arg Asp Leu Ile Trp Asp Tyr Ile Met Arg Thr Leu 1505 1510 1515 1520
- Arg Asp Thr Ser His Ser Ile Leu Lys Val Leu Ser Asn Ala Leu Ser 1525 1530 1535
- His Pro Lys Val Phe Lys Arg Phe Trp Asp Cys Gly Val Leu Asn Pro 1540 1545 1550
- Ile Tyr Gly Pro Asn Ile Ala Ser Gln Asp Gln Ile Lys Leu Ala Leu 1555 1560 1565

- 271 -

Ser	Ile 1570	-	Glu	Tyr	Ser	Leu 1575	_	Leu	Phe	Met	<b>Ar</b> g 1580		Trp	Leu	Asn
Gly 1585		Ser	Leu	Glu	11e 1590	_	Ile	Cys	Asp	Ser 1595	_	Met	Glu	Val	Ala 1600
Asn	Asp	λrg	Lys	Gln 1605		Phe	Ile	Ser	<b>A</b> rg 1610		Leu	Ser	Phe	Val 1615	_
Сув	Leu	Ala	Glu 1620		Ala	Ser	Phe	Gly 1625		Asn	Leu	Leu	Asn 1630		Thr
Tyr	Leu	Glu 1635	Arg	Leu	Asp	Leu	Leu 1640		Gln	Tyr	Leu	Glu 1645		Asn	Ile
Lys	Glu 1650	_	Pro	Thr	Leu	Lys 1655	-	Val	Gln	Ile	Ser 1660	_	Leu	Leu	Ile
Lys 1665		Phe	Pro	Ser	Thr 1670		Thr	Tyr	Val	Arg 1675	_	Thr	Ala	Ile	Lys 1680
Tyr	Leu	Arg	Ile	Arg 1685	_	Ile	Ser	Pro	Pro 1690		Val	Ile	Asp	Авр 1695	•
Asp	Pro	Val	Glu 1700	_	Glu	Asn	Met	Leu 1705	-	Asn	Ile	Val	Lys 1710		Ile
Asn	Asp	Asn 1715	Cys ;	Asn	Lys	Asp	Asn 1720	-	Gly	Asn	Lys	Ile 1725		Asn	Phe
Trp	Gly 1730		Ala	Leu	Lys	Asn 1735	_	Gln	Val	Leu	Lys 1740		Arg	Ser	Ile
Thr 1745		Авр	Ser	yab	Asp 1750		Asp	Arg	Leu	Asp 1755		Asn	Thr	Ser	Gly 1760
Leu	Thr	Leu	Pro	Gln 1765	-	Gly	Asn	Tyr	Leu 1770		His	Gln	Leu	Arg 1775	
Phe	Gly	Ile	Asn 1780		Thr	Ser	Сув	Leu 1785	_	Ala	Leu	Glu	Leu 1790		Gln
Ile	Leu	Met 1795	Lys	Glu	Val	Asn	Lys 1800	_	Lys	Asp	Arg	Leu 1805		Leu	Gly
Glu	Gly 1810		Gly	Ala	Met	Leu 1815		Сув	Tyr	Asp	Ala 1820		Leu	Gly	Pro
Ala 1825		Asn	Tyr	Tyr	<b>Asn</b> 1830		Gly	Leu	Asn	Ile 1835		Asp	Val	Ile	Gly 1840

- 272 -

				184	5				100						
									185	,0				185	55
Lye	Let ,	ı Gly	186	Val	Thr	Gln	Ile	186	ı Asr	Arg	Val	Lys	Val 187		Phe
Asn	Gly	187	Pro	Asn	Ser	Thr	188	Ile 80	Gly	' Asn	Met	Glu 188		Glu	Ser
Leu	189	Trp	Ser	Glu	Leu	Asn 189	. <b>А</b> вр 5	Lys	Ser	Ile	Gly 190		Val	His	Сув
Asp 190	Met 5	Glu	Gly	Ala	Ile 191	Gly 0	Lys	Ser	Glu	Glu 191		Val	Leu	His	Glu 1920
His	Tyr	Ser	Val	Ile 192	Arg 5	Ile	Thr	Tyr	Leu 193		Gly	Asp	Asp	Asp 193	Val 5
Val	Leu	Val	Ser 194	Lys 0	Ile	Ile	Pro	Thr 194	Ile 5	Thr	Pro	Asn	Trp		Arg
Ile	Leu	Tyr 195	Leu 5	Tyr	Lys	Leu	Tyr 196	Trp 0	Lys	Asp	Val	Ser 196		Ile	Ser
Leu	Lys 197	Thr 0	Ser	Asn	Pro	Ala 197	Ser 5	Thr	Glu	Leu	Tyr 1980		Ile	Ser	Lys
Asp 198	Ala 5	Tyr	Сув	Thr	Ile 1990	Met )	Glu	Pro	Ser	Glu 1995		Val	Leu	Ser	Lys 2000
Leu	Lys	Arg	Leu	Ser 2005	Leu 5	Leu	Glu	Glu	Asn 2010		Leu	Leu	Lys	Trp 2015	
Ile	Leu	Ser	Lys 2020	Lys	Arg	Asn	Asn	Glu 202	Trp	Leu	His	His	Glu 2030		Lys
Glu	Gly	Glu 2035	Arg	Asp	Tyr	Gly	Ile 2040	Met )	Arg	Pro		His 2045		Ala	Leu
Gln	Ile 2050	Phe	Gly	Phe	Gln	Ile 2055	Asn	Leu	Asn		Leu 2060		Lys	Glu	Phe
Leu 2065	Ser	Thr	Pro	Asp	Leu 2070	Thr	Asn	Ile	Asn	Asn 2075	Ile	Ile	Gln		Phe 2080
Gln	Arg	Thr	Ile	Lys 2085	Asp	Val	Leu	Phe	Glu 2090	Trp	Ile	Asn		Thr 2095	
Asp	qaA	Lys	Arg 2100	His	Lys	Leu	Gly	Gly 2105	Arg	Tyr i	Asn :		Phe 2110		Leu

2125

Lys Asn Lys Gly Lys Leu Arg Leu Ser Arg Arg Leu Val Leu Ser

2120

2115

- 273 -

Trp Ile Ser Leu Ser Leu Ser Thr Arg Leu Leu Thr Gly Arg Phe Pro 2130 2140

Asp Glu Lys Phe Glu His Arg Ala Gln Thr Gly Tyr Val Ser Leu Ala 2145 2150 2155 2160

Asp Thr Asp Leu Glu Ser Leu Lys Leu Ser Lys Asn Ile Ile Lys 2165 2170 2175

Asn Tyr Arg Glu Cys Ile Gly Ser Ile Ser Tyr Trp Phe Leu Thr Lys 2180 2185 2190

Glu Val Lys Ile Leu Met Lys Leu Ile Gly Gly Ala Lys Leu Leu Gly 2195 2200 2205

Ile Pro Arg Gln Tyr Lys Glu Pro Glu Asp Gln Leu Leu Glu Asn Tyr 2210 2215 2220

Asn Gln His Asp Glu Phe Asp Ile Asp 2225 2230

- (2) INFORMATION FOR SEQ ID NO:21:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 15462 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: RNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACCARACAAG AGAAGAAACT TGCTTGGTAA TATAAATTTA ACTTAAAATT AACTTAGGAT 60 TTANGACATT GACTAGAAGG TCAAGAAAAG GGAACTCTAT AATTTCAAAA ATGTTGAGCC 120 TATTTGATAC ATTTAATGCA CGTAGGCAAG AAAACATAAC AAAATCAGCC GGTGGAGCTA 180 TCATTCCTGG ACAGAAAAAT ACTGTCTCTA TATTCGCCCT TGGACCGACA ATAACTGATG 240 ATAATGAGAA AATGACATTA GCTCTTCTAT TTCTATCTCA TTCACTAGAT AATGAGAAAC 300 AACATGCACA AAGGGCAGGG TTCTTGGTGT CTTTATTGTC AATGGCTTAT GCCAATCCAG 360 AGCTCTACCT AACAACAAAT GGAAGTAATG CAGATGCCAA GTATGTCATA TACATGATTG 420 AGAAAGATCT AAAACGGCAA AAGTATGGAG GATTTGTGGT TAAGACGAGA GAGATGATAT 480

- 274 -

ATGAAAAGAC	AACTGATTGG	ATATTTGGAA	GTGACCTGGA	TTATGATCAG	GAAACTATGT	540
TGCAGAACGG	CAGGAACAAT	TCAACAATTG	AAGACCTTGT	CCACACATTT	GGGTATCCAT	600
CATGTTTAGG	AGCTCTTATA	ATACAGATCT	GGATAGTTCT	GGTCAAAGCT	ATCACTAGTA	660
TCTCAGGGTT	AAGAAAAGGC	·TTTTTCACCC	GATTGGAAGC	TTTCAGACAA	GATGGAACAG	720
TGCAGGCAGG	GCTGGTATTG	AGCGGTGACA	CAGTGGATCA	GATTGGGTCA	ATCATGCGGT	780
CTCAACAGAG	CTTGGTAACT	CTTATGGTTG	AAACATTAAT	AACAATGAAT	ACCAGCAGAA	840
ATGACCTCAC	AACCATAGAA	AAGAATATAC	AAATTGTTGG	CAACTACATA	AGAGATGCAG	900
GTCTCGCTTC	ATTCTTCAAT	ACAATCAGAT	ATGGAATTGA	GACCAGAATG	GCAGCTTTGA	960
CTCTATCCAC	TCTCAGACCA	GATATCAATA	GATTAAAAGC	TTTGATGGAA	CTGTATTTAT	1020
CAAAGGGACC	ACGCGCTCCT	TTCATCTGTA	TCCTCAGAGA	TCCTATACAT	GGTGAGTTCG	1080
CACCAGGCAA	CTATCCTGCC	ATATGGAGCT	ATGCAATGGG	GGTGGCAGTT	GTACAAAATA	1140
GAGCCATGCA	ACAGTATGTG	ACGGGAAGAT	CATATCTAGA	CATTGATATG	TTCCAGCTAG	1200
GACAAGCAGT	AGCACGTGAT	GCCGAAGCTC	AAATGAGCTC	AACACTGGAA	GATGAACTTG	1260
GAGTGACACA	CGAAGCTAAA	GAAAGCTTGA	AGAGACATAT	AAGGAACATA	AACAGTTCAG	1320
AGACATCTTT	CCACAAACCG	ACAGGTGGAT	CAGCCATAGA	GATGGCAATA	GATGAAGAGC	1380
CAGAACAATT	CGAACATAGA	GCAGATCAAG	AACAAAATGG	AGAACCTCAA	TCATCCATAA	1440
TTCAATATGC	CTGGGCAGAA	GGAAATAGAA	GCGATGATCA	GACTGAGCAA	GCTACAGAAT	1500
CTGACAATAT	CAAGACCGAA	CAACAAAACA	TCAGAGACAG	ACTAAACAAG	AGACTCAACG	1560
ACAAGAAGAA	ACAAAGCAGT	CAACCACCCA	CTAATCCCAC	AAACAGAACA	AACCAGGACG	1620
AAATAGATGA	TCTGTTTAAC	GCATTTGGAA	GCAACTAATC	GAATCAACAT	TTTAATCTAA	1680
ATCAATAATA	aataagaaaa	ACTTAGGATT	AAAGAATCCT	ATCATACCGG	AATATAGGGT	1740
GGTAAATTTA	GAGTCTGCTT	GAAACTCAAT	CAATAGAGAG	TTGATGGAAA	GCGATGCTAA	1800
AAACTATCAA	ATCATGGATT	CTTGGGAAGA	GGAATCAAGA	GATAAATCAA	CTAATATCTC	1860
CTCGGCCCTC	AACATCATTG	AATTCATACT	CAGCACCGAC	CCCCAAGAAG	ACTTATCGGA	1920
AAACGACACA	ATCAACACAA	GAACCCAGCA	ACTCAGTGCC	ACCATCTGTC	AACCAGAAAT	1980
CAAACCAACA	GAAACAAGTG	agaaagatag	TGGATCAACT	GACAAAAATA	GACAGTCTGG	2040

- 275 -

GTCATCACAC	GAATGTACAA	CAGAAGCAAA	AGATAGAAAC	ATTGATCAGG	AAACTGTACA	2100
GAGAGGACCT	GGGAGAAGAA	GCAGCTCAGA	TAGTAGAGCT	GAGACTGTGG	TCTCTGGAGG	2160
AATCCCCAGA	AGCATCACAG	ATTCTAAAAA	TGGAACCCAA	AACACGGAGG	ATATTGATCT	2220
CAATGAAATT	AGAAAGATGG	ATAAGGACTC	TATTGAGGGG	AAAATGCGAC	AATCTGCAAA	2280
TGTTCCAAGC	GAGATATCAG	GAAGTGATGA	CATATTTACA	ACAGAACAAA	GTAGAAACAG	2340
TGATCATGGA	AGAAGCCTGG	AATCTATCAG	TACACCTGAT	ACAAGATCAA	TAAGTGTTGT	2400
TACTGCTGCA	ACACCAGATG	ATGAAGAAGA	AATACTAATG	AAAAATAGTA	GGACAAAGAA	2460
AAGTTCTTCA	ACACATCAAG	AAGATGACAA	AAGAATTAAA	AAAGGGGGAA	AAGGGAAAGA	2520
CTGGTTTAAG	AAATCAAAAG	ATACCGACAA	CCAGATACCA	ACATCAGACT	ACAGATCCAC	2580
ATCAAAAGGG	CAGAAGAAAA	TCTCAAAGAC	AACAACCACC	AACACCGACA	CAAAGGGGCA	2640
AACAGAAATA	CAGACAGAAT	CATCAGAAAC	ACAATCCTCA	TCATGGAATC	TCATCATCGA	2700
CAACAACACC	GACCGGAACG	AACAGACAAG	CACAACTCCT	CCAACAACAA	CTTCCAGATC	2760
AACTTATACA	AAAGAATCGA	TCCGAACAAA	CTCTGAATCC	AAACCCAAGA	CACAAAAGAC	2820
AAATGGAAAG	GAAAGGAAGG	ATACAGAAGA	GAGCAATCGA	TTTACAGAGA	GGGCAATTAC	2880
TCTATTGCAG	AATCTTGGTG	TAATTCAATC	CACATCAAAA	CTAGATTTAT	ATCAAGACAA	2940
ACGAGTTGTA	TGTGTAGCAA	ATGTACTAAA	CAATGTAGAT	ACTGCATCAA	AGATAGATTT	3000
CCTGGCAGGA	TTAGTCATAG	GGGTTTCAAT	GGACAACGAC	ACAAAATTAA	CACAGATACA	3060
AAATGAAATG	CTAAACCTCA	AAGCAGATCT	AAAGAAAATG	GACGAATCAC	ATAGAAGATT	3120
GATAGAAAAT	CAAAGAGAAC	AACTGTCATT	GATCACGTCA	CTAATTTCAA	ATCTCAAAAT	3180
		AGAAAGACCA				3240
		AAAAGATCAA				3300
					CACTAGAGAA	3360
					CAACAAGACT	3420
					ACAACAGCAA	3480
					AAAATGATGA	3540
AGAAGTATCT	GAATTAATGG	ACATGTTCAA	TGAAGATGTC	AACAATTGCC	AATGATCCAA	3600

- 276 -

	•
CAAAGAAACG ACACCGAACA AACAGACAAG AAACAACAGT AGATCAAAAC CTGTCAACAC	3660
ACACAAAATC AAGCAGAATG AAACAACAGA TATCAATCAA TATACAAATA AGAAAAACTT	3720
AGGATTAAAG AATAAATTAA TCCTTGTCCA AAATGAGTAT AACTAACTCT GCAATATACA	3780
CATTCCCAGA ATCATCATTC TCTGAAAATG GTCATATAGA ACCATTACCA CTCAAAGTCA	3840
ATGAACAGAG GAAAGCAGTA CCCCACATTA GAGTTGCCAA GATCGGAAAT CCACCAAAAC	3900
ACGGATCCCG GTATTTAGAT GTCTTCTTAC TCGGCTTCTT CGAGATGGAA CGAATCAAAG	3960
ACAAATACGG GAGTGTGAAT GATCTCGACA GTGACCCGAG TTACAAAGTT TGTGGCTCTG	4020
GATCATTACC AATCGGATTG GCTAAGTACA CTGGGAATGA CCAGGAATTG TTACAAGCCG	4080
CAACCAAACT GGATATAGAA GTGAGAAGAA CAGTCAAAGC GAAAGAGATG GTTGTTTACA	4140
CGGTACAAAA TATAAAACCA GAACTGTACC CATGGTCCAA TAGACTAAGA AAAGGAATGC	4200
TGTTCGATGC CAACAAAGTT GCTCTTGCTC CTCAATGTCT TCCACTAGAT AGGAGCATAA	4260
AATTTAGAGT AATCTTCGTG AATTGTACGG CAATTGGATC AATAACCTTG TTCAAAATTC	4320
CTAAGTCAAT GGCATCACTA TCTCTAACCA ACACAATATC AATCAATCTG CAGGTACACA	4380
TAAAAACAGG GGTTCAGACT GATTCTAAAG GGATAGTTCA AATTTTGGAT GAGAAAGGCG	4440
AAAAATCACT GAATTTCATG GTCCATCTCG GATTGATCAA AAGAAAAGTA GGCAGAATGT	4500
ACTCTGTTGA ATACTGTAAA CAGAAAATCG AGAAAATGAG ATTGATATTT TCTTTAGGAC	4560
TAGTTGGAGG AATCAGTCTT CATGTCAATG CAACTGGGTC CATATCAAAA ACACTAGCAA	4620
GTCAGCTGGT ATTCAAAAGA GAGATTTGTT ATCCTTTAAT GGATCTAAAT CCGCATCTCA	4680
ATCTAGTTAT CTGGGCTTCA TCAGTAGAGA TTACAAGAGT GGATGCAATT TTCCAACCTT	4740
CTTTACCTGG CGAGTTCAGA TACTATCCTA ATATTATTGC AAAAGGAGTT GGGAAAATCA	4800
AACAATGGAA CTAGTAATCT CTATTTAGT CCGGACGTAT CTATTAAGCC GAAGCAAATA	4860
AAGGATAATC AAAAACTTAG GACAAAAGAG GTCAATACCA ACAACTATTA GCAGTCACAC	4920
TCGCAAGAAT AAGAGAGAAG GGACCAAAAA AGTCAAATAG GAGAAATCAA AACAAAAGGT	4980
ACAGAACACC AGAACAACAA AATCAAAACA TCCAACTCAC TCAAAACAAA AATTCCAAAA	5040
GAGACCGGCA ACACAACAAG CACTGAACAC AATGCCAACT TCAATACTGC TAATTATTAC	5100
AACCATGATC ATGGCATCTT TCTGCCAAAT AGATATCACA AAACTACAGC ACGTAGGTGT	5160

- 277 -

ATTGGTCAA	C AGTCCCAAAG	GGATGAAGAT	T ATCACÁRARO	TTTGAAACAI	GATATCTAAT	5220
TTTGAGCCT	: ATACCAAAA	TAGAAGACTO	TAACTCTTGT	GGTGACCAA	AGATCAAGCA	5280
ATACAAGAA	TTATTGGATA	GACTGATCAT	CCCTTTATAT	GATGGATTA	GATTACAGAA	5340
AGATGTGATA	GTAACCAATC	AAGAATCCAA	TGAAAACACT	GATCCCAGA	CAAAACGATT	5400
CTTTGGAGG	GTAATTGGAA	CCATTGCTCT	GGGAGTAGCA	ACCTCAGCAC	AAATTACAGC	5460
GGCAGTTGCT	CTGGTTGAAG	CCAAGCAGGC	AAGATCAGAC	ATCGAAAAAC	TCAAAGAAGC	5520
AATTAGGGAC	ACAAATAAAG	CAGTGCAGTC	AGTTCAGAGC	TCCATAGGAA	ATTTAATAGT	5580
AGCAATTAAA	TCAGTCCAGG	ATTATGTTAA	CAAAGAAATC	GTGCCATCGA	TTGCGAGGCT	5640
					ACTCAGAATT	5700
					AATTACAAGG	5760
					CAGTTGATAA	5820
					TAGATGTTGA	5880
					GGCTGCTGAA	5940
	TACAAAGTAG					6000
	AGCCATATCA					6060
	GCATTCAGCA					6120
	AGCTGCTTAT					6180
	CCAAGATATG					6240
	TGCAACGGAA					6300
	CATAAAGAAT					6360
	ACTCTTGCAT				TAGAAGAATC	6420
					TAGAAGAATC  GGCATCAATC	6480
					GGCATCAATC TTAATATAAC	6540
					TTAATATAAC TGGATCAAAA	6600
					ATATTAAAAT	6660
					VIVIIVAVI	6720

- 278 -

TATAAAAAAC	TTAGGAGTAA	AGTTACGCAA	TCCAACTCTA	CTCATATAAT	TGAGGAAGGA	6780
CCCAATAGAC	AAATCCAAAT	TCGAGATGGA	ATACTGGAAG	CATACCAATC	ACGGAAAGGA	6840
TGCTGGCAAT	GAGCTGGAGA	CGTCTATGGC	TACTCATGGC	AACAAGCTCA	CTAATAAGAT	6900
AATATACATA	TTATGGACAA	TAATCCTGGT	GTTATTATCA	ATAGTCTTCA	TCATAGTGCT	6960
AATTAATTCC	ATCAAAAGTG	AAAAGGCCCA	CGAATCATTG	CTGCAAGACA	TAAATAATGA	7020
GTTTATGGAA	ATTACAGAAA	AGATCCAAAT	GGCATCGGAT	AATACCAATG	ATCTAATACA	7080
GTCAGGAGTG	AATACAAGGC	TTCTTACAAT	TCAGAGTCAT	GTCCAGAATT	ACATACCAAT	7140
ATCATTGACA	CAACAGATGT	CAGATCTTAG	GAAATTCATT	AGTGAAATTA	CAATTAGAAA	7200
TGATAATCAA	GAAGTGCTGC	CACAAAGAAT	AACACATGAT	GTAGGTATAA	AACCTTTAAA	7260
TCCAGATGAT	TTTTGGAGAT	GCACGTCTGG	TCTTCCATCT	TTAATGAAAA	CTCCAAAAAT	7320
AAGGTTAATG	CCAGGGCCGG	GATTATTAGC	TATGCCAACG	ACTGTTGATG	GCTGTGTTAG	7380
AACTCCGTCT	TTAGTTATAA	ATGATCTGAT	TTATGCTTAT	ACCTCAAATC	TAATTACTCG	7440
AGGTTGTCAG	GATATAGGAA	AATCATATCA	AGTCTTACAG	ATAGGGATAA	TAACTGTAAA	7500
CTCAGACTTG	GTACCTGACT	TAAATCCTAG	GATCTCTCAT	ACCTTTAACA	TAAATGACAA	7560
TAGGAAGTCA	TGTTCTCTAG	CACTCCTAAA	TACAGATGTA	TATCAACTGT	GTTCAACTCC	7620
CAAAGTTGAT	GAAAGATCAG	ATTATGCATC	ATCAGGCATA	GAAGATATTG	TACTTGATAT	7680
TGTCAATTAT	GATGGTTCAA	TCTCAACAAC	AAGATTTAAG	AATAATAACA	TAAGCTTTGA	7740
TCAACCATAT	GCTGCACTAT	ACCCATCTGT	TGGACCAGGG	ATATACTACA	AAGGCAAAAT	7800
AATATTTCTC	GGGTATGGAG	GTCTTGAACA	TCCAATAAAT	GAGAATGTAA	TCTGCAACAC	7860
AACTGGGTGC	CCCGGGAAAA	CACAGAGAGA	CTGTAATCAA	GCGTCTCATA	GTCCATGGTT	7920
TTCAGATAGG	AGGATGGTCA	ACTCCATCAT	TGTTGCTGAC	AAAGGCTTAA	ACTCAATTCC	7980
AAAATTGAAA	GTATGGACGA	TATCTATGCG	ACAAAATTAC	TGGGGGTCAG	AAGGAAGGTT	8040
ACTTCTACTA	GGTAACAAGA	TCTATATATA	TACAAGATCT	ACAAGTTGGC	ATAGCAAGTT	8100
ACAATTAGGA	ATAATTGATA	TTACTGATTA	CAGTGATATA	AGGATAAAAT	GGACATGGCA	8160
TAATGTGCTA	TCAAGACCAG	GAAACAATGA	ATGTCCATGG	GGACATTCAT	GTCCAGATGG	8220
ATGTATAACA	GGAGTATATA	CTGATGCATA	TCCACTCAAT	CCCACAGGGA	GCATTGTGTC	8280

- 279 -

ATCTGTCATA	TTAGACTCAC	AAAAATCGAG	AGTGAACCCA	GTCATAACTT	ACTCAACAGC	8340
AACCGAAAGA	GTAAACGAGC	TGGCCATCCT	AAACAGAACA	CTCTCAGCTG	GATATACAAC	8400
AACAAGCTGC	ATTACACACT	ATAACAAAGG	ATATTGTTTT	CATATAGTAG	AAATAAATCA	8460
TAAAAGCTTA	AACACATTTC	AACCCATGTT	GTTCAAAACA	GAGATTCCAA	AAAGCTGCAG	8520
TTAATCATAA	TTAACCATAA	TATGCATCAA	TCTATCTATA	ATACAAGTAT	ATGATAAGTA	8580
ATCAGCAATC	AGACAATAGA	CAAAAGGGAA	ATATAAAAA	CTTAGGAGCA	AAGCGTGCTC	8640
GGGAAATGGA	CACTGAATCT	AACAATGGCA	CTGTATCTGA	CATACTCTAT	CCTGAGTGTC	8700
ACCTTAACTC	TCCTATCGTT	AAAGGTAAAA	TAGCACAATT	ACACACTATT	ATGAGTCTAC	8760
CTCAGCCTTA	TGATATGGAT	GACGACTCAA	TACTAGTTAT	CACTAGACAG	AAAATAAAAC	8820
TTAATAAATT	GGATAAAGA	CAACGATCTA	TTAGAAGATT	AAAATTAATA	TTAACTGAAA	8880
AAGTGAATGA	CTTAGGAAAA	TACACATTTA	TCAGATATCC	AGAAATGTCA	AAAGAAATGT	8940
TCARATTATA	TATACCTGGT	ATTAACAGTA	AAGTGACTGA	ATTATTACTT	AAAGCAGATA	9000
GAACATATAG	TCAAATGACT	GATGGATTAA	GAGATCTATG	GATTAATGTG	CTATCAAAAT	9060
TAGCCTCAAA	AAATGATGGA	AGCAATTATG	ATCTTAATGA	AGAAATTAAT	AATATATCGA	9120
AAGTTCACAC	AACCTATAAA	TCAGATAAAT	GGTATAATCC	ATTCAAAACA	TGGTTTACTA	9180
TCAAGTATGA	TATGAGAAGA	TTACAAAAAG	CTCGAAATGA	GATCACTTTT	AATGTTGGGA	9240
AGGATTATAA	CTTGTTAGAA	GACCAGAAGA	ATTTCTTATT	GATACATCCA	GAATTGGTTT	9300
TGATATTAGA	TAAACAAAAC	TACAATGGTT	ATCTAATTAC	TCCTGAATTA	GTATTGATGT	9360
ATTGTGACGT	AGTCGAAGGC	CGATGGAATA	TAAGTGCATG	TGCTAAGTTA	GATCCAAAAT	9420
TACAATCTAT	GTATCAGAAA	GGTAATAACC	TGTGGGAAGT	GATAGATAAA	TTGTTTCCAA	9480
TTATGGGAGA	AAAGACATTT	GATGTGATAT	CGTTATTAGA	ACCACTTGCA	TTATCCTTAA	9540
TTCAAACTCA	TGATCCTGTT	AAACAACTAA	GAGGAGCTTT	TTTAAATCAT	GTGTTATCCG	9600
AGATGGAATT	AATATTTGAA	TCTAGAGAAT	CGATTAAGGA	ATTTCTGAGT	GTAGATTACA	9660
TTGATAAAAT	TTTAGATATA	TTTAATAAGT	CTACAATAGA	TGAAATAGCA	GAGATTTTCT	9720
CTTTTTTAG	AACATTTGGG	CATCCTCCAT	TAGAAGCTAG	TATTGCAGCA	GAAAAGGTTA	9780
GAAAATATAT	GTATATTGGA	AAACAATTAA	AATTTGACAC	TATTAATAAA	TGTCATGCTA	9840

- 280 -

TCTTCTGTAC AATAATAATT AACGGATATA GAGAGAGGCA TGGTGGACAG TGGCCTCCTG	9900
TGACATTACC TGATCATGCA CACGAATTCA TCATAAATGC TTACGGTTCA AACTCTGCGA	9960
TATCATATGA GAATGCTGTT GATTATTACC AGAGCTTTAT AGGAATAAAA TTCAATAAAT	10020
TCATAGAGCC TCAGTTAGAT GAGGATTTGA CAATTTATAT GAAAGATAAA GCATTATCTC	10080
CARARANTC ARRITGGGAC ACAGITTATC CIGCATCIAR ITTACIGIAC CGIACIARCG	10140
CATCCAACGA ATCACGAAGA TTAGTTGAAG TATTTATAGC AGATAGTAAA TTTGATCCTC	10200
ATCAGATATT GGATTATGTA GAATCTGGGG ACTGGTTAGA TGATCCAGAA TTTAATATTT	10260
CTTATAGTCT TAAAGAAAAA GAGATCAAAC AGGAAGGTAG ACTCTTTGCA AAAATGACAT	10320
ACAAAATGAG AGCTACACAA GTTTTATCAG AGACACTACT TGCAAATAAC ATAGGAAAAT	10380
TCTTTCAAGA AAATGGGATG GTGAAGGGAG AGATTGAATT ACTTAAGAGA TTAACAACCA	10440
TATCAATATC AGGAGTTCCA CGGTATAATG AAGTGTACAA TAATTCTAAA AGCCATACAG	10500
ATGACCTTAA AACCTACAAT AAAATAAGTA ATCTTAATTT GTCTTCTAAT CAGAAATCAA	10560
AGAAATTTGA ATTCAAGTCA ACGGATATCT ACAATGATGG ATACGAGACT GTGAGCTGTT	10620
TCCTAACAAC AGATCTCAAA AAATACTGTC TTAATTGGAG ATATGAATCA ACAGCTCTAT	10680
TTGGAGAAAC TTGCAACCAA ATATTTGGAT TAAATAAATT GTTTAATTGG TTACACCCTC	10740
GTCTTGAAGG AAGTACAATC TATGTAGGTG ATCCTTACTG TCCTCCATCA GATAAAGAAC	10800
ATATATCATT AGAGGATCAC CCTGATTCTG GTTTTTACGT TCATAACCCA AGAGGGGGTA	10860
TAGAAGGATT TTGTCAAAAA TTATGGACAC TCATATCTAT AAGTGCAATA CATCTAGCAG	10920
CTGTTAGAAT AGGCGTGAGG GTGACTGCAA TGGTTCAAGG AGACAATCAA GCTATAGCTG	10980
TAACCACAAG AGTACCCAAC AATTATGACT ACAGAGTTAA GAAGGAGATA GTTTATAAAG	11040
ATGTAGTGAG ATTTTTTGAT TCATTAAGAG AAGTGATGGA TGATCTAGGT CATGAACTTA	11100
AATTAAATGA AACGATTATA AGTAGCAAGA TGTTCATATA TAGCAAAAGA ATCTATTATG	11160
ATGGGAGAAT TCTTCCTCAA GCTCTAAAAG CATTATCTAG ATGTGTCTTC TGGTCAGAGA	11220
CAGTAATAGA CGAAACAAGA TCAGCATCTT CAAATTTGGC AACATCATTT GCAAAAGCAA	11280
TTGAGAATGG TTATTCACCT GTTCTAGGAT ATGCATGCTC AATTTTTAAG AACATTCAAC	11340
AACTATATAT TGCCCTTGGG ATGAATATCA ATCCAACTAT AACACAGAAT ATCAGAGATC	11400

- 281 -

AGTATTTAG	GAATCCAAA	T TGGATGCAA	T ATGCCTCTT	T AATACCTGC	T AGTGTTGGGG	11460
GATTCAATCA	CATGGCCAT	G TCAAGATGT	T TTGTAAGGA	A TATTGGTGA	T CCATCAGTTG	11520
CCGCATTGGC	TGATATTAA	A AGATTTATT	A AGGCGAATC	I ATTAGACCG	A AGTGTTCTTT	11580
ATAGGATTAT	GAATCAAGA	A CCAGGTGAG	I CATCITITI	I TGACTGGGC	T TCAGATCCAT	11640
ATTCATGCAA	TTTACCACAL	TCTCAAAAT	A TAACCACCA	r Gataaaaa	T ATAACAGCAA	11700
GGAATGTATT	ACAAGATTC	CCAAATCCAT	TATTATCTG	ATTATTCAC	A AATACAATGA	11760
TAGAAGAAGA	TGAAGAATTI	A GCTGAGTTC	TGATGGACA	GAAGGTAAT:	T CTCCCTAGAG	11820
TTGCACATGA	TATTCTAGAT	AATTCTCTCA	CAGGAATTA	AAATGCCAT	A GCTGGAATGT	11880
TAGATACGAC	AAAATCACTA	ATTCGGGTTG	GCATAAATAG	AGGAGGACTO	ACATATAGTT	11940
TGTTGAGGAA	AATCAGTAA1	TACGATCTAG	; TACAATATG	AACACTAAG	AGGACTTTGC	12000
GACTAATTGT	AAGTGATAAA	ATCAAGTATG	AAGATATGTG	TTCGGTAGAC	CTTGCCATAG	12060
CATTGCGACA	AAAGATGTGG	ATTCATTTAT	CAGGAGGAAG	GATGATAAGT	GGACTTGAAA	12120
CGCCTGACCC	ATTAGAATTA	CTATCTGGGG	TAGTAATAAC	AGGATCAGAA	CATTGTAAAA	12180
TATGTTATTC	TTCAGATGGC	ACAAACCCAT	ATACTTGGAT	GTATTTACCO	GGTAATATCA	12240
AAATAGGATC	AGCAGAAACA	GGTATATCGT	CATTAAGAGT	TCCTTATTTT	GGATCAGTCA	12300
CTGATGAAAG	ATCTGAAGCA	CAATTAGGAT	ATATCAAGAA	TCTTAGTAAA	CCTGCAAAAG	12360
CCGCAATAAG	AATAGCAATG	ATATATACAT	GGGCATTTGG	TAATGATGAG	ATATCTTGGA	12420
TGGAAGCCTC	ACAGATAGCA	CAAACACGTG	CAAATTTTAC	ACTAGATAGT	CTCAAAATTT	12480
TAACACCGGT	AGCTACATCA	ACAAATTTAT	CACACAGATT	TAAGGATACT	GCAACTCAGA	12540
TGAAATTCTC	CAGTACATCA	TTGATCAGAG	TCAGCAGATT	TATAACAATG	TCCAATGATA	12600
ACATGTCTAT	Caaagaagct	AATGAAACCA	AAGATACTAA	TCTTATTTAT	CAACAAATAA	12660
TGTTAACAGG	Attaagtgtt	TTCGAATATT	TATTTAGATT	AAAAGAAACC	ACAGGACACA	12720
ACCCTATAGT	TATGCATCTG	CACATAGAAG	ATGAGTGTTG	TATTAAAGAA	AGTTTTAATG	12780
ATGAACATAT	TAATCCAGAG	TCTACATTAG	AATTAATTCG	ATATCCTGAA	AGTAATGAAT	12840
TTATTTATGA	FAAAGACCCA	CTCAAAGATG	TGGACTTATC	AAAACTTATG	GTTATTAAAG	12900
ACCATTCTTA (	CACAATTGAT	ATGAATTATT	GGGATGATAC	TGACATCATA	CATGCAATTT	12960

- 282 -

CAATATGTAC	TGCAATTACA	ATAGCAGATA	CTATGTCACA	ATTAGATCGA	GATAATTTAA	13020
AAGAGATAAT	AGTTATTGCA	AATGATGATG	ATATTAATAG	CTTAATCACT	GAATTTTTGA	13080
CTCTTGACAT	ACTTGTATTT	CTCAAGACAT	TTGGTGGATT	attagtaaat	CAATTTGCAT	13140
ACACTCTTTA	TAGTCTAAAA	ATAGAAGGTA	GGGATCTCAT	TTGGGATTAT	ATAATGAGAA	13200
CACTGAGAGA	TACTTCCCAT	TCAATATTAA	AAGTATTATC	TAATGCATTA	TCTCATCCTA	13260
AAGTATTCAA	GAGGTTCTGG	GATTGTGGAG	TTTTAAACCC	TATTTATGGT	CCTAATATTG	13320
CTAGTCAAGA	CCAGATAAA	CTTGCCCTAT	CTATATGTGA	ATATTCACTA	GATCTATTTA	13380
TGAGAGAATG	GTTGAATGGT	GTATCACTTG	AAATATACAT	TTGTGACAGC	GATATGGAAG	13440
TTGCAAATGA	TAGGAAACAA	GCCTTTATTT	CTAGACACCT	TTCATTTGTT	TGTTGTTTAG	13500
CAGAAATTGC	ATCTTTCGGA	CCTAACCTGT	TAAACTTAAC	ATACTTGGAG	AGACTTGATC	13560
TATTGAAACA	ATATCTTGAA	TTAAATATTA	AAGAAGACCC	TACTCTTAAA	TATGTACAAA	13620
TATCTGGATT	AAATTAATTA	TCGTTCCCAT	CAACTGTAAC	ATACGTAAGA	AAGACTGCAA	13680
TCAAATATCT	AAGGATTCGC	GGTATTAGTC	CACCTGAGGT	AATTGATGAT	TGGGATCCGG	13740
TAGAAGATGA	AAATATGCTG	GATAACATTG	TCAAAACTAT	AAATGATAAC	TGTAATAAAG	13800
ATAATAAAGG	GAATAAAATT	AACAATTTCT	GGGGACTAGC	ACTTAAGAAC	TATCAAGTCC	13860
TTAAAATCAG	ATCTATAACA	AGTGATTCTG	ATGATAATGA	TAGACTAGAT	GCTAATACAA	13920
GTGGTTTGAC	ACTTCCTCAA	GGAGGGAATT	ATCTATCGCA	. TCAATTGAGA	TTATTCGGAA	13980
TCAACAGCAC	TAGTTGTCTG	AAAGCTCTTG	AGTTATCACA	AATTTTAATG	AAGGAAGTCA	14040
					GCATGTTATG	14100
					ACAGATGTAA	14160
					DATTAAAAAA*	14220
					CCTAATTCAA	
					GATAAGTCCA	
					ACTGTTCTAC	
					GTTGTTTAG	
TTTCCAAAA	TATACCTAC	A ATCACTCCG	A ATTGGTCTAC	AATACTTTA!	CTATATAAAT	14520

- 283 -

TATATTG	GAA	AGATGTAAG	T ATAATATCA	C TCAAAACTTC	TAATCCTGCA	TCAACAGAAT	1458
TATATCT.	AAT	TTCGAAAGAT	GCATATTGT:	CTATAATGGA	ACCTAGTGAA	ATTGTTTAT	1464
CAAAACT	TAA	AAGATTGTCA	CTCTTGGAAG	AAAATAATCT	ATTAAAATGG	ATCATTTAT	1470
CAAAGAA	GAG	Gaataatgaa	TGGTTACATO	ATGAAATCAA	AGAAGGAGAA	AGAGATTATG	14760
GAATCAT	gag	ACCATATCAT	ATGGCACTAC	AAATCTTTGG	ATTTCAAATC	AATTTAAATC	14820
ATCTGGC	GAA	AGAATTTTTA	TCAACCCCAG	ATCTGACTAA	TATCAACAAT	ATAATCCAAA	14880
GTTTTCAC	3CG	AACAATAAAG	GATGTTTAT	TTGAATGGAT	TAATATAACT	CATGATGATA	14940
AGAGACAT	AAT	ATTAGGCGGA	AGATATAACA	TATTCCCACT	GAAAAATAAG	GGAAAGTTAA	15000
GACTGCTA	ATC	GAGAAGACTA	GTATTAAGTT	GGATTTCATT	ATCATTATCG	ACTCGATTAC	15060
TTACAGG1	rcg	CTTTCCTGAT	GAAAAATTTG	AACATAGAGC	ACAGACTGGA	TATGTATCAT	15120
ragctgat	AC	TGATTTAGAA	TCATTAAAGT	TATTGTCGAA	AAACATCATT	AAGAATTACA	15180
BAGAGTGT	AT	AGGATCAATA	TCATATTGGT	TTCTAACCAA	AGAAGTTAAA	ATACTTATGA	15240
lattgatt	.GG	TGGTGCTAAA	TTATTAGGAA	TTCCCAGACA	ATATAAAGAA	CCCGAAGACC	15300
GTTATTA.	GA .	AAACTACAAT	CAACATGATG	AATTTGATAT	CGATTAAAAC	ATAAATACAA	15360
'GAAGATA	TA '	TCCTAACCTT	TATCTTTAAG	CCTAGGAATA	GACAAAAAGT	AAGAAAAACA	15420
GTAATAT.	AT	ATATACCAAA	CAGAGTTCTT	CTCTTGTTTG	GT		15462

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2233 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:
- Met Asp Thr Glu Ser Asn Asn Gly Thr Val Ser Asp Ile Leu Tyr Pro
- Glu Cys His Leu Asn Ser Pro Ile Val Lys Gly Lys Ile Ala Gln Leu

- 284 -

			20					25			•		30		
His	Thr	11e 35	Met	Ser	Leu	Pro	Gln 40	Pro	Tyr	Asp	Met	Авр 45	Asp	qaA	Ser
Ile	Leu 50	Val	Ile	Thr	Arg	Gln 55	Lys	Ile	Lys	Leu	Asn 60	Lya	Leu	Двр	Lys
Arg 65	Gln	Arg	Ser	Ile	Arg 70	Arg	Leu	Lys	Leu	Ile 75	Leu	Thr	Glu	Lys	Val BO
Asn	Asp	Leu	Gly	Lys 85	Tyr	Thr	Phe	Ile	Arg 90	Tyr	Pro	Glu	Met	Ser 95	Lys
Glu	Met	Phe	Lys 100	Leu	Tyr	Ile	Pro	Gly 105	Ile	Asn	Ser	Lys	Val 110	Thr	Glu
Leu	Leu	Leu 115	Lys	Ala	Asp	Arg	Thr 120	Tyr	Ser	Gln	Met	Thr 125	qaA	Gly	Leu
Arg	Asp 130	Leu	Trp	Ile	Asn	Val 135	Leu	Ser	Lys	Leu	Ala 140	Ser	Lys	Asn	Авр
Gly 145	Ser	Asn	Tyr	Asp	Leu 150	Asn	Glu	Glu	Ile	Asn 155	Asn	Ile	Ser	Lys	Val 160
His	Thr	Thr	Tyr	Lys 165	Ser	Asp	Lys	Trp	Tyr 170	Asn	Pro	Phe	Lys	Thr 175	Trp
Phe	Thr	Ile	Lys 180	Tyr	Asp	Met	Arg	<b>Arg</b> 185	Leu	Gln	Lys	Ala	Arg 190	Asn	Glu
Ile	Thr	Phe 195	Asn	Val	Gly	Lys	Авр 200	Tyr	Asn	Leu	Leu	Glu 205	Asp	Gln	Lys
Asn	Phe 210	Leu	Leu	lle	His	Pro 215	Glu	Leu	Val	Leu	11e 220	Leu	qaA	Lys	Gln
Asn 225	Tyr	Asn	Gly	Tyr	Leu 230	Ile	Thr	Pro	Glu	Leu 235	Val	Leu	Met	Tyr	Cys 240
Asp	Val	Val	Glu	Gly 245	Arg	Trp	Asn	Ile	Ser 250	Ala	Сув	Ala	Lys	Leu 255	Asp
Pro	Lys	Leu	Gln 260	Ser	Met	Tyr	Gln	Lув 265	Gly	Asn	Asn	Leu	Trp 270	Glu	Val
Ile	Asp	<b>L</b> ув 275	Leu	Phe	Pro	Ile	Met 280	Gly	Glu	Lys	Thr	Phe 285	Asp	Val	Ile
Ser	Leu 290	Leu	Glu	Pro	Leu	Ala 295	Leu	Ser	Leu	Ile	Gln 300	Thr	His	Asp	Pro

- 285 -

Va 30	l Ly 5	s Gl	n Le	u Ar	g Gl <sub>)</sub> 310	Ala D	a Ph	e Le	u As	n Hi 31		l Le	u Se	r Gl	u Mei 320	
Gl	n Fe	u Il	e Ph	e Gl:	u Ser	Arq	g Gl	u Se	7 Il		s Gl	u Ph	e Le	u Se 33	r Val	L
Ası	р Ту:	r Il	e As;	p Lys	s Ile	Let	ı Ası	345		e Ası	n Ly	s Se	r Th		e Asp	>
Gli	ı Ile	35!	a Gli 5	u Ile	Phe	Sez	Phe 360		Arg	g Th	r Ph	e Gl; 36		s Pr	o Pro	)
Lev	370	a Ala	a Sez	r Ile	Ala	Ala 375	Glu	ı Lys	Va]	l Arg	380		r Me	t Ty	r Ile	t
Gly 385	Lys	Glr	ı Let	ı Lys	Phe 390	Asp	Thr	Ile	Asn	1 Lys		Hi:	Ala	a Ile	Phe 400	
Сув	Thr	Ile	Ile	11e 405	Asn	Gly	Tyr	Arg	Glu 410		His	Gly	Gl <sub>y</sub>	Glr 415	Trp	
Pro	Pro	Val	Thr 420	Leu	Pro	Asp	His	Ala 425	His	Glu	Phe	Ile	11e		Ala	
Tyr	Gly	Ser 435	Asn	Ser	Ala	Ile	Ser 440	Tyr	Glu	Asn	Ala	Val		Tyr	Tyr	
Gln	Ser 450	Phe	Ile	Gly	Ile	Lув 455	Phe	Asn	Lys	Phe	Ile 460		Pro	Gln	Leu	
Авр 465	Glu	Asp	Leu	Thr	Ile 470	Tyr	Met	Lys	Asp	Lys 475	Ala	Leu	Ser	Pro	Lys 480	
Lys	Ser	Asn	Trp	Asp 485	Thr	Val	Tyr	Pro	Ala 490	Ser	Asn	Leu	Leu	Tyr 495	Arg	
Thr	Asn	Ala	Ser 500	Asn	Glu	Ser	Arg	Arg 505	Leu	Val	Glu	Val	Phe 510	Ile	Ala	
Asp	Ser	Lys 515	Phe	Asp	Pro	His	Gln 520	Ile	Leu	Asp	Tyr	Val 525	Glu	Ser	Gly	
Asp	Trp 530	Leu	Asp	Asp	Pro	Glu 535	Phe	Asn	Ile	Ser	Tyr 540	Ser	Leu	Lys	Glu	
Lув 545	Glu	Ile	Lys	Gln	Glu ( 550	Gly .	Arg	Leu		Ala 555	Lys	Met	Thr	Tyr	Lys 560	
Met	Arg	Ala	Thr	Gln 565	Val 1	Leu :	Ser		Thr 570	Leu	Leu	Ala		Asn 575	Ile	

- 286 -

Gly	Lys	Phe	Phe 580	Gln	Glu	Asn	Gly	Met 585	Val	Lys	Gly	Glu	Ile 590	Glu	Leu
Leu	Lys	Arg 595	Leu	Thr	Thr	Ile	Ser 600	Ile	Ser	Gly	Val	Pro 605	Arg	Tyr	Asn
Glu	Val 610	Tyr	Asn	Asn	Ser	Lys 615	Ser	His	Thr	Авр	Asp 620	Leu	Lys	Thr	Tyr
Asn 625	Lys	Ile	Ser	Asn	Leu 630	Asn	Leu	Ser	Ser	Asn 635	Gln	Lys	Ser	Lys	Lys 640
Phe	Glu	Phe	Lys	Ser 645	Thr	Авр	Ile	Tyr	Asn 650	Asp	Gly	Tyr	Glu	Thr 655	Val
Ser	Сув	Phe	Leu 660	Thr	Thr	Asp	Leu	Lys 665	Lys	Tyr	Сув	Leu	Asn 670	Trp	Arg
Tyr	Glu	Ser 675	Thr	Ala	Leu	Phe	Gly 680	Glu	Thr	Сув	Asn	Gln 685	Ile	Phe	Gly
Leu	Asn 690	Lys	Leu	Phe	Asn	Trp 695	Leu	His	Pro	Arg	Leu 700	Glu	Gly	Ser	Thr
Ile 705	Tyr	Val	Gly	Asp	Pro 710	Tyr	Суя	Pro	Pro	Ser 715	Asp	Lys	Glu	His	11e 720
Ser	Leu	Glu	Asp	H15 725	Pro	Asp	Ser	Gly	Phe 730	Tyr	Val	His	Asn	Pro 735	Arg
Gly	Gly	Ile	Glu 740	Gly	Phe	Сув	Gln	Lys 745	Leu	Trp	Thr	Leu	11e 750	Ser	Ile
Ser	Ala	11e 755	His	Leu	Ala	Ala	Val 760	Arg	Ile	Gly	Val	Arg 765	Val	Thr	Ala
Met	Val 770	Gln	Gly	Asp	Asn	Gln 775	Ala	Ile	Ala	Val	Thr 780	Thr	Arg	Val	Pro
Asn 785	Asn	Tyr	Asp	Tyr	Arg 790	Val	Lys	Lys	Glu	11e 795	Val	Tyr	Lys	Asp	Val 800
Val	Arg	Phe	Phe	<b>Авр</b> 805	Ser	Leu	Arg	Glu	Val 810	Met	Asp	Asp	Leu	Gly 815	His
Glu	Leu	Lys	Leu 820	Asn	Glu	Thr	Ile	Ile 825	Ser	Ser	Lys	Met	Phe 830	Ile	Tyr
Ser	Lys	Arg 835	Ile	Tyr	Tyr	Asp	Gly 840	Arg	Ile	Leu	Pro	Gln 845	Ala	Leu	Lув
Ala	Leu	Ser	Arg	Сув	Val	Phe	Trp	Ser	Glu	Thr	Val	Ile	Asp	Glu	Thr

- 287 -

	85	0				85	5				86	)			
86:	g Se 5	r Al	a Se	r Se	870	ı Let	ı Ala	a Th	r Se	r Phe 875		Lys	Ala	Ile	Glu 880
Ası	n Gl	у Ту	r Se	r Pro	o Val	Let	ı Gly	y Ty:	r Al 89	а Сує 0	3 Sei	Ile	Phe	Lys 895	
Ile	e Gl	n Gl	n Le	1 Ty:	r Ile	Ala	. Let	90!	y Me S	t Asn	ıle	Asn	Pro 910		Ile
Thi	Gl:	n Ası 91!	n Ile 5	a Arç	y Asp	Glr	920	Phe	a Ar	g Asn	Pro	Asn 925		Met	Gln
Тут	930	a Sei	r Lei	ı Ile	Pro	Ala 935	Ser	Va]	G1;	y Gly	Phe 940		His	Met	Ala
Me t 945	Sei	Arg	у Суғ	Phe	950	Arg	Asn	Ile	Gly	7 Asp 955		Ser	Val	Ala	Ala 960
Leu	Ala	Asg	Ile	Lya 965	Arg	Phe	Ile	Lys	970	Asn	Leu	Leu	Asp	Arg 975	Ser
Val	Leu	Туг	980	Ile	Met	Asn	Gln	Glu 985	Pro	Gly	Glu	Ser	Ser 990	Phe	Phe
Asp	Trp	995	Ser	Asp	Pro	Tyr	Ser 100	Сув 0	Asn	Leu	Pro	Gln 1005		Gln	Asn
Ile	Thr 101	Thr 0	Met	Ile	Lys	Asn 101	Ile 5	Thr	Ala	Arg	Asn 102		Leu	Gln	Asp
Ser 102	Pro 5	Asn	Pro	Leu	Leu 103(	Ser )	Gly	Leu	Phe	Thr 1035		Thr	Met		Glu 1040
Glu	Asp	Glu	Glu	Leu 104	Ala 5	Glu	Phe	Leu	Met 105	Asp 0	Arg	Lys		Ile 1055	
Pro	Arg	Val	Ala 106	His )	Asp	Ile	Leu	Asp 106!	Asn 5	Ser	Leu		Gly 1070		Arg
Asn	Ala	Ile 107	Ala 5	Gly	Met	Leu	Asp 1080	Thr	Thr	Lys	Ser	Leu 1085		Arg '	Val
Gly	11e 109	Asn )	Arg	Gly	Gly	Leu 1095	Thr	Tyr	Ser	Leu	Leu 1100		Lys :	Ile :	Ber
Asn 1105	Tyr	Asp	Leu	Val	Gln 1110	Tyr	Glu	Thr	Leu	Ser .		Thr :	Leu 1		Jeu 1120
Ile	Val	Ser	Asp	Lys 1125	Ile	Lys	Tyr	Glu	Asp 1130	Met	Сув	Ser '		Asp I 1135	eu

- 288 -

- Ala Ile Ala Leu Arg Gln Lys Met Trp Ile His Leu Ser Gly Gly Arg 1140 1145 1150
- Met Ile Ser Gly Leu Glu Thr Pro Asp Pro Leu Glu Leu Leu Ser Gly 1155 1160 1165
- Val Val Ile Thr Gly Ser Glu His Cys Lys Ile Cys Tyr Ser Ser Asp 1170 1175 1180
- Gly Thr Asn Pro Tyr Thr Trp Met Tyr Leu Pro Gly Asn Ile Lys Ile 1185 1190 1195 1200
- Gly Ser Ala Glu Thr Gly Ile Ser Ser Leu Arg Val Pro Tyr Phe Gly 1205 1210 1215
- Ser Val Thr Asp Glu Arg Ser Glu Ala Gln Leu Gly Tyr Ile Lys Asn 1220 1225 1230
- Leu Ser Lys Pro Ala Lys Ala Ala Ile Arg Ile Ala Met Ile Tyr Thr 1235 1240 1245
- Trp Ala Phe Gly Asn Asp Glu Ile Ser Trp Met Glu Ala Ser Gln Ile 1250 1255 1260
- Ala Gln Thr Arg Ala Asn Phe Thr Leu Asp Ser Leu Lys Ile Leu Thr 1265 1270 1275 1280
- Pro Val Ala Thr Ser Thr Asn Leu Ser His Arg Phe Lys Asp Thr Ala 1285 1290 1295
- Thr Gln Met Lys Phe Ser Ser Thr Ser Leu Ile Arg Val Ser Arg Phe 1300 1305 1310
- Ile Thr Met Ser Asn Asp Asn Met Ser Ile Lys Glu Ala Asn Glu Thr 1315 1320 1325
- Lys Asp Thr Asn Leu Ile Tyr Gln Gln Ile Met Leu Thr Gly Leu Ser 1330 1335 1340
- Val Phe Glu Tyr Leu Phe Arg Leu Lys Glu Thr Thr Gly His Asn Pro 1345 1350 1355 1360
- Ile Val Met His Leu His Ile Glu Asp Glu Cys Cys Ile Lys Glu Ser 1365 1370 1375
- Phe Asn Asp Glu His Ile Asn Pro Glu Ser Thr Leu Glu Leu Ile Arg 1380 1385 1390
- Tyr Pro Glu Ser Asn Glu Phe Ile Tyr Asp Lys Asp Pro Leu Lys Asp 1395 1400 1405

- 289 -

- Val Asp Leu Ser Lys Leu Met Val Ile Lys Asp His Ser Tyr Thr Ile 1410 1415 1420
- Asp Met Asn Tyr Trp Asp Asp Thr Asp Ile Ile His Ala Ile Ser Ile 1425 1430 1435 1440
- Cys Thr Ala Ile Thr Ile Ala Asp Thr Met Ser Gln Leu Asp Arg Asp 1445 1450 1455
- Asn Leu Lys Glu Ile Ile Val Ile Ala Asn Asp Asp Asp Ile Asn Ser 1460 1465 1470
- Leu Ile Thr Glu Phe Leu Thr Leu Asp Ile Leu Val Phe Leu Lys Thr 1475 1480 1485
- Phe Gly Gly Leu Leu Val Asn Gln Phe Ala Tyr Thr Leu Tyr Ser Leu 1490 1495 1500
- Lys Ile Glu Gly Arg Asp Leu Ile Trp Asp Tyr Ile Met Arg Thr Leu 1505 1510 1515 1520
- Arg Asp Thr Ser His Ser Ile Leu Lys Val Leu Ser Asn Ala Leu Ser 1525 1530 1535
- His Pro Lys Val Phe Lys Arg Phe Trp Asp Cys Gly Val Leu Asn Pro 1540 1545 1550
- Ile Tyr Gly Pro Asn Ile Ala Ser Gln Asp Gln Ile Lys Leu Ala Leu 1555 1560 1565
- Ser Ile Cys Glu Tyr Ser Leu Asp Leu Phe Met Arg Glu Trp Leu Asn 1570 1575 1580
- Gly Val Ser Leu Glu Ile Tyr Ile Cys Asp Ser Asp Met Glu Val Ala 1585 1590 1595 1600
- Asn Asp Arg Lys Gln Ala Phe Ile Ser Arg His Leu Ser Phe Val Cys 1605 1610 1615
- Cys Leu Ala Glu Ile Ala Ser Phe Gly Pro Asn Leu Leu Asn Leu Thr 1620 1625 1630
- Tyr Leu Glu Arg Leu Asp Leu Leu Lys Gln Tyr Leu Glu Leu Asn Ile 1635 1640 1645
- Lys Glu Asp Pro Thr Leu Lys Tyr Val Gln Ile Ser Gly Leu Leu Ile 1650 1660
- Lys Ser Phe Pro Ser Thr Val Thr Tyr Val Arg Lys Thr Ala Ile Lys 1665 1670 1675 1680
- Tyr Leu Arg Ile Arg Gly Ile Ser Pro Pro Glu Val Ile Asp Asp Trp

- 290 -

				1685	5				1690	)		٠		1699	5
Asp	Pro	Val	Glu 1700		Glu	Asn	Met	Leu 1705	_	Asn	Ile	Val	Lys 1710		Ile
Asn	Asp	Asn 171		Asn	Lys	Asp	Asn 1720		Gly	Asn	Lys	Ile 1725		Asn	Phe
Trp	Gly 1730		Ala	Leu	Lys	Asn 1735		Gln	Val	Leu	Lys 1740		Arg	Ser	Ile
Thr 174		Asp	Ser	Asp	Asp 1750		Asp	Arg	Leu	Asp 1755		Asn	Thr	Ser	Gly 1760
Leu	Thr	Leu	Pro	Gln 1769		Gly	Asn	Tyr	Leu 1770		His	Gln	Leu	Arg 1775	
Phe	Gly	Ile	Asn 1780		Thr	Ser	Сув	Leu 1785		Ala	Leu	Glu	Leu 1790		Gln
Ile	Leu	Met 1795	_	Glu	Val	Asn	Lys 1800	_	Lys	Asp	Arg	Leu 1805		Leu	Gly
Glu	Gly 1810		Gly	Ala	Met	Leu 1815		Сув	Tyr	Asp	Ala 1820		Leu	Gly	Pro
Ala 182		Asn	Tyr	Tyr	Asn 1830		Gly	Leu	Asn	Ile 1835		Asp	Val	Ile	Gly 1840
Gln	Arg	Glu	Leu	Lys 1845		Phe	Pro	Ser	Glu 1850		Ser	Leu	Val	Gly 1855	_
Lys	Leu	Gly	Asn 1860		Thr	Gln	Ile	Leu 1865		Arg	Val	Lys	Val 1870		Phe
Asn	Gly	Asn 1875		Asn	Ser	Thr	Trp 1880		Gly	Asn	Met	Glu 1885	_	Glu	Ser
Leu	Ile 1890	_	Ser	Glu	Leu	Asn 1895	_	Lys	Ser	Ile	Gly 1900		Val	His	Сув
Авр 190		Glu	Gly	Ala	Ile 1910	_	Lys	Ser	Glu	Glu 1915		Val	Leu	His	Glu 1920
His															
	Tyr	Ser	Val	Ile 1925		Ile	Thr	Tyr	Leu 193(		Gly	Asp	Авр	Asp 1935	
	Tyr			1925 Lys	5				1930	)			-	1935 Ser	5

- 291 -

- Leu Lys Thr Ser Asn Pro Ala Ser Thr Glu Leu Tyr Leu Ile Ser Lys 1970 1975 1980
- Asp Ala Tyr Cys Thr Ile Met Glu Pro Ser Glu Ile Val Leu Ser Lys 1985 1990 1995 2000
- Leu Lys Arg Leu Ser Leu Leu Glu Glu Asn Asn Leu Leu Lys Trp Ile 2005 2010 2015
- Ile Leu Ser Lys Lys Arg Asn Asn Glu Trp Leu His His Glu Ile Lys
  2020 2025 2030
- Glu Gly Glu Arg Asp Tyr Gly Ile Met Arg Pro Tyr His Met Ala Leu 2035 2040 2045
- Gln Ile Phe Gly Phe Gln Ile Asn Leu Asn His Leu Ala Lys Glu Phe 2050 2055 2060
- Leu Ser Thr Pro Asp Leu Thr Asn Ile Asn Asn Ile Ile Gln Ser Phe 2065 2070 2075 2080
- Gln Arg Thr Ile Lys Asp Val Leu Phe Glu Trp Ile Asn Ile Thr His 2085 2090 2095
- Asp Asp Lys Arg His Lys Leu Gly Gly Arg Tyr Asn Ile Phe Pro Leu 2100 2105 2110
- Lys Asn Lys Gly Lys Leu Arg Leu Leu Ser Arg Arg Leu Val Leu Ser 2115 2120 2125
- Trp Ile Ser Leu Ser Leu Ser Thr Arg Leu Leu Thr Gly Arg Phe Pro 2130 2135 2140
- Asp Glu Lys Phe Glu His Arg Ala Gln Thr Gly Tyr Val Ser Leu Ala 2145 2150 2155 2160
- Asp Thr Asp Leu Glu Ser Leu Lys Leu Leu Ser Lys Asn Ile Ile Lys 2165 2170 2175
- Asn Tyr Arg Glu Cys Ile Gly Ser Ile Ser Tyr Trp Phe Leu Thr Lys 2180 2185 2190
- Glu Val Lys Ile Leu Met Lys Leu Ile Gly Gly Ala Lys Leu Leu Gly 2195 2200 2205
- Ile Pro Arg Gln Tyr Lys Glu Pro Glu Asp Gln Leu Leu Glu Asn Tyr 2210 2215 2220
- Asn Gln His Asp Glu Phe Asp Ile Asp 2225 2230

- 292 -

# (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS: "
  - (A) LENGTH: 15218 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA (genomic)

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ACGCGAAA	AA	ATGCGTACTA	CAAACTTGCA	CATTCGAAAA	AAATGGGGCA	AATAAGAACT	60
TGATAAGT	3C	TATTTAAGTC	TAACCTTTTC	AATCAGAAAT	GGGGTGCAAT	TCACTGAGCA	120
TGATAAAG	3T	TAGATTACAA	AATTTATTTG	ACAATGACGA	AGTAGCATTG	TTAAAAATAA	180
CATGTTAT	AC	TGATAAATTA	ATTCTTCTGA	CCAATGCATT	AGCCAAAGCA	GCAATACATA	240
CAATTAAA!	ГT	AAACGGCATA	GTTTTTATAC	ATGTTATAAC	AAGCAGTGAA	GTGTGCCCTG	300
ATAACAAT	ΑT	TGTAGTGAAA	TCTAACTTTA	CAACAATGCC	AATACTACAA	AATGGAGGAT	360
ACATATGG	GA	ATTGATTGAG	TTGACACACT	GCTCTCAATT	AAACGGTTTA	ATGGATGATA	420
ATTGTGAA	AT	CAAATTTTCT	AAAAGACTAA	GTGACTCAGT	AATGACTAAT	TATATGAATC	480
AAATATCT	ga	CTTACTTGGG	CTTGATCTCA	ATTCATGAAT	TATGTTTAGT	CTAATTCAAT	540
AGACATGT	GT	TTATTACCAT	TTTAGTTAAT	ATAAAAACTC	ATCAAAGGGA	AATGGGGCAA	600
ATAAACTC	AC	CTAATCAATC	AAACCATGAG	CACTACAAAT	GACAACACTA	CTATGCAAAG	660
ATTGATGA	TC	ACAGACATGA	GACCCCTGTC	AATGGATTCA	ATAATAACAT	CTCTTACCAA	720
AGAAATCA	TC	ACACACAAAT	TCATATACTT	GATAAACAAT	GAATGTATTG	TAAGAAAACT	780
TGATGAAA	GA	CAAGCTACAT	TTACATTCTT	AGTCAATTAT	GAGATGAAGC	TACTGCACAA	840
AGTAGGGA	GT	ACCAAATACA	AAAAATACAC	TGAATATAAT	ACAAAATATG	GCACTTTCCC	900
CATGCCTA	TA	TTTATCAATC	ACGGCGGGTT	TCTAGAATGT	ATTGGCATTA	AGCCTACAAA	960
ACACACTO	CT	ATAATATACA	AATATGACCT	CAACCCGTGA	ATTCCAACAA	AAAAACCAAC	1020
CCAACCAA	AC	CAAACTATTC	CTCAAACAAC	AGTGCTCAAT	agttaagaag	GAGCTAATCC	1080
ATTTTAGT	'AA	AATAAAATT	AAGTAAAGCC	AATAACATAA	ATTGGGGCAA	ATACAAAGAT	1140

- 293 -

GGCTCTTAGC AAAGTCAAGT TGAATGATAC ATTAAATAAG GATCAGCTGC TGTCATCCAG	1200
CAAATACACT ATTCAACGTA GTACAGGAGA TAATATTGAC ACTCCCAATT ATGATGTGCA	1260
AAAACACCTA AACAAACTAT GTGGTATGCT ATTAATCACT GAAGATGCAA ATCATAAATT	1320
CACAGGATTA ATAGGTATGT TATATGCTAT GTCCAGGTTA GGAAGGGAAG	1380
GATACTTAAA GATGCTGGAT ATCATGTTAA AGCTAATGGA GTAGATATAA CAACATATCG	1440
TCAAGATATA AATGGAAAGG AAATGAAATT CGAAGTATTA ACATTATCAA GCTTGACATC	1500
AGAAATACAA GTCAATATTG AGATAGAATC TAGAAAGTCC TACAAAAAA TGCTAAAAGA	1560
GATGGGAGAA GTGGCTCCAG AATATAGGCA TGATTCTCCA GACTGTGGGA TGATAATACT	1620
GTGTATAGCT GCACTTGTGA TAACCAAATT AGCAGCAGGA GACAGATCAG GTCTTACAGC	1680
AGTAATTAGG AGGGCAAACA ATGTCTTAAA AAACGAAATA AAACGATACA AGGGCCTCAT	1740
ACCAAAGGAT ATAGCTAACA GTTTTTATGA AGTGTTTGAA AAACACCCTC ATCTTATAGA	1800
TGTTTTCGTG CACTTTGGCA TTGCACAATC ATCCACAAGA GGGGGTAGTA GAGTTGAAGG	1860
AATCTTTGCA GGATTGTTTA TGAATGCCTA TGGTTCAGGG CAAGTAATGC TAAGATGGGG	1920
AGTTTTAGCC AAATCTGTAA AAAATATCAT GCTAGGACAT GCTAGTGTCC AGGCAGAAAT	1980
GGAGCAAGTT GTGGAAGTCT ATGAGTATGC ACAGAAGTTG GGAGGAGAAG CTGGATTCTA	2040
CCATATATTG AACAATCCAA AAGCATCATT GCTGTCATTA ACTCAATTTC CCAACTTCTC	2100
AAGTGTGGTC CTAGGCAATG CAGCAGGTCT AGGCATAATG GGAGAGTATA GAGGTACACC	2160
AAGAAACCAG GATCTTTATG ATGCAGCTAA AGCATATGCA GAGCAACTCA AAGAAAATGG	2220
AGTAATAAAC TACAGTGTAT TAGACTTAAC AGCAGAAGAA TTGGAAGCCA TAAAGCATCA	2280
ACTCAACCCC AAAGAAGATG ATGTAGAGCT TTAAGTTAAC AAAAAATACG GGGCAAATAA	2340
GTCAACATGG AGAAGTTTGC ACCTGAATTT CATGGAGAAG ATGCAAATAA CAAAGCTACC	2400
AAATTCCTAG AATCAATAAA GGGCAAGTTC GCATCATCCA AAGATCCTAA GAAGAAAGAT	2460
AGCATAATAT CTGTTAACTC AATAGATATA GAAGTAACTA AAGAGAGCCC GATAACATCT	2520
GGCACCAACA TCATCAATCC AACAAGTGAA GCCGACAGTA CCCCAGAAAC AAAAGCCAAC	2580
TACCCAAGAA AACCCCTAGT AAGCTTCAAA GAAGATCTCA CCCCAAGTGA CAACCCTTTT	2640
TCTAAGTTGT ACAAGGAAAC AATAGAAACA TTTGATAACA ATGAAGAAGA ATCTAGCTAC	2700

- 294 -

•	TCATATGAAG	AGATAAATGA	TCAAACAAAT	GACAACATTA	CAGCAAGACT	AGATAGAATT	2760
•	TAAAAADTAE	TAAGTGAAAT	ATTAGGAATG	CTCCATACAT	TAGTAGTTGC	aagtgcagga	2820
•	CCCACTTCAG	CTCGCGATGG	AATAAGAGAT	GCTATGGTTG	GTCTAAGAGA	AGAGATGATA	2880
(	GAAAAAATAA	GAGCGGAAGC	ATTAATGACC	AATGATAGGT	TAGAGGCTAT	GGCAAGACTT	2940
	AGGAATGAGG	AAAGCGAAAA	AATGGCAAAA	GACACCTCAG	ATGAAGTGTC	TCTTAATCCA	3000
•	ACTTCCAAAA	AATTGAGTGA	CTTGTTGGAA	GACAACGATA	GTGACAATGA	TCTATCACTT	3060
•	GATGATTTT	GATCAGCGAT	CAACTCACTC	AGCAATCAAC	AACATCAATA	AAACAGACAT	3120
	CAATCCATTG	AATCAACTGC	CAGACCGAAC	AAACAAACGT	CCATCAGTAG	AACCACCAAC	3180
	CAATCAATCA	ACCAATTGAT	CAATCAGCAA	CCCGACAAAA	TTAACAATAT	AGTAACAAAA	3240
	AAAGAACAAG	ATGGGGCAAA	TATGGAAACA	TACGTGAACA	AGCTTCACGA	AGGCTCCACA	3300
	TACACAGCAG	CTGTTCAGTA	CAATGTTCTA	GAAAAAGATG	ATGATCCTGC	ATCACTAACA	3360
	ATATGGGTGC	CTATGTTCCA	GTCATCTGTG	CCAGCAGACT	TGCTCATAAA	AGAACTTGCA	3420
	agcatcaa <b>t</b> a	TACTAGTGAA	GCAGATCTCT	ACGCCCAAAG	GACCTTCACT	ACGAGTCACG	3480
	ATTAACTCAA	GAAGTGCTGT	GCTGGCTCAA	ATGCCTAGTA	ATTTCATCAT	AAGCGCAAAT	3540
	GTATCATTAG	atgaaagaag	CAAATTAGCA	TATGATGTAA	CTACACCTTG	TGAAATCAAA	3600
	GCATGCAGTC	TAACATGCTT	AAAAGTAAAA	AGTATGTTAA	CTACAGTCAA	AGATCTTACC	3660
	ATGAAGACAT	TCAACCCCAC	TCATGAGATC	ATTGCTCTAT	GTGAATTTGA	AAATATTATG	3720
	ACATCAAAAA	GAGTAATAAT	ACCAACCTAT	CTAAGATCAA	TTAGTGTCAA	GAACAAGGAT	3780
	CTGAACTCAC	TAGAAAATAT	AGCAACCACC	GAATTCAAAA	ATGCTATCAC	CAATGCAAAA	3840
	ATTATTCCTT	ATGCAGGATT	AGTGTTAGTT	ATCACAGTTA	CTGACAATAA	AGGAGCATTC	3900
	AAATATATCA	AACCACAGAG	TCAATTTATA	GTAGATCTTG	GTGCCTACCT	AGAAAAAGAG	3960
	AGCATATATT	ATGTGACTAC	TAATTGGAAG	CATACAGCTA	CACGTTTTTC	AATCAAACCA	4020
	CTAGAGGATT	AAACTTAATT	ATCAACACTG	AATGACAGGT	CCACATATAT	CCTCAAACTA	4080
	CACACTATAT	CCARACATCA	TAAACATCTA	CACTACACAC	TTCATCACAC	AAACCAATCC	4140
	CACTCAAAAT	CCAAAATCAC	TACCAGCCAC	TATCTGCTAG	ACCTAGAGTG	CGAATAGGTA	4200
	AATAAAACCA	AAATATGGGG	TAAATAGACA	TTAGTTAGAG	TTCAATCAAT	CTTAACAACC	4260

- 295 -

ATTTATACC	G CCAATTCAA	CACATATACT	A TAAATCTTAI	AATGGGAAA	r ACATCCATCA	4320
CAATAGAAT	T CACAAGCAA	A TTTTGGCCC	T ATTTTACACT	T AATACATAT	ATCTTAACTC	4380
TAATCTTTT	r actaattati	A ATCACTATT!	A TGATTGCAAT	ACTARATAR	G CTAAGTGAAC	4440
ATAAAGCAT:	I CTGTAACAA	ACTCTTGAAC	TAGGACAGAT	GTATCAAATC	AACACATAGA	4500
GTTCTACCA	r Tatgetgtgt	CAAATTATA	1 TCCTGTATAT	ATAAACAAA	AAATCCAATC	4560
TTCTCACAG	A GTCATGGTG1	CGCAAAACCA	CGCTAACTAI	CATGGTAGCA	TAGAGTAGTT	4620
ATTTAAAAA	TAACATAATO	ATGAATTGTT	' AGTATGAGAT	CAAAAACAAC	: ATTGGGGCAA	4680
ATGCAACCAT	GTCCAAACAC	AAGAATCAAC	GCACTGCCAG	GACTCTAGAA	AAGACCTGGG	4740
ATACTCTTA	TCATCTAATI	GTAATATCCT	CTTGTTTATA	CAGATTAAAT	TTARARTCTA	4800
TAGCACAAAT	AGCACTATCA	GTTTTGGCAA	TGATAATCTC	AACCTCTCTC	ATAATTGCAG	4860
CCATAATATT	CATCATCTCT	GCCAATCACA	AAGTTACACT	AACAACGGTC	ACAGTTCAAA	4920
CARTARARA	CCACACTGAA	AAAAACATCA	CCACCTACCC	TACTCAAGTC	TCACCAGAAA	4980
GGGTTAGTTC	ATCCAAGCAA	CCCACAACCA	CATCACCAAT	CCACACAAGT	TCAGCTACAA	5040
CATCACCCAA	TACAAAATCA	GAAACACACC	ATACAACAGC	ACAAACCAAA	GGCAGAACCA	5100
CCACTTCAAC	ACAGACCAAC	AAGCCAAGCA	CAAAACCACG	TCCAAAAAAT	CCACCAAAAA	5160
AAGATGATTA	CCATTTTGAA	GTGTTCAACT	TCGTTCCCTG	CAGTATATGT	GGCAACAATC	5220
AACTTTGCAA	ATCCATCTGC	AAAACAATAC	CAAGCAACAA	ACCAAAGAAG	AAACCAACCA	5280
TCAAACCCAC	AAACAAACCA	ACCACCAAAA	CCACAAACAA	AAGAGACCCA	AAAACACCAG	5340
CCAAAACGAC	Gaaaaagaa	ACTACCACCA	ACCCAACAAA	AAAACTAACC	CTCAAGACCA	5400
CAGAAAGAGA	CACCAGCACC	TCACAATCCA	CTGCACTCGA	CACAACCACA	TTAAAACACA	5460
CAGTCCAACA	GCAATCCCTC	CTCTCAACCA	CCCCGAAAA	CACACCCAAC	TCCACACAAA	5520
CACCCACAGC	ATCCGAGCCC	TCCACACCAA	ACTCCACCCA	AAAAACCCAG	CCACATGCTT	5580
AGTTATTCAA	AAACTACATC	TTAGCAGAGA	ACCGTGATCT	ATCAAGCAAG	AACGAAATTA	5640
AACCTGGGGC	AAATAACCAT	GGAGTTGATG	ATCCACAAGT	CAAGTGCAAT	CTTCCTAACT	5700
CTTGCTATTA	ATGCATTGTA	CCTCACCTCA	AGTCAGAACA	TAACTGAGGA	GTTTTACCAA	5760
TCGACATGTA	GTGCAGTTAG	CAGAGGTTAT	TTTAGTGCTT	TAAGAACAGG	TTGGTATACT	5820

- 296 -

AGTGTCATAA	CAATAGAATT	AAGTAATATA	AAAGAAACCA	AATGCAATGG	AACTGACACT	5880
AAAGTAAAAC	TTATGAAACA	AGAATTAGAT	AAGTATAAGA	ATGCAGTAAC	AGAATTACAG	5940
CTACTTATGC	AAAACACACC	AGCTGTCAAC	AACCGGGCCA	GAAGAGAAGC	ACCACAGTAT	6000
ATGAACTACA	CAATCAATAC	CACTAAAAAC	CTAAATGTAT	CAATAAGCAA	GAAGAGGAAA	6060
CGAAGATTTC	TAGGCTTCTT	GTTAGGTGTG	GGATCTGCAA	TAGCAAGTGG	TATAGCTGTA	6120
TCAAAAGTTC	TACACCTTGA	AGGAGAAGTG	AACAAGATCA	AAAATGCTTT	GTTGTCTACA	6180
AACAAAGCTG	TAGTCAGTTT	ATCAAATGGG	GTCAGTGTTT	TAACCAGCAA	agtgttagat	6240
CTCAAGAATT	ACATAAATAA	CCAATTATTA	CCCATAGTAA	ATCAACAGAG	CTGTCGCATC	6300
TCCAACATTG	AAACAGTTAT	AGAATTCCAG	CAGAAGAACA	GCAGATTGTT	GGAAATCACC	6360
AGAGAATTTA	GTGTCAATGC	AGGTGTAACA	ACACCTTTAA	GCACTTACAT	GTTGACAAAC	6420
AGTGAGTTAC	TATCATTAAT	CAATGATATG	CCTATAACAA	ATGATCAGAA	AAAATTAATG	6480
TCAAGCAATG	TTCAGATAGT	AAGGCAACAA	AGTTATTCCA	TCATGTCTAT	AATAAAGGAA	6540
GAAGTCCTTG	CATATGTTGT	ACAGCTGCCT	ATCTATGGTG	TAATAGATAC	ACCTTGCTGG	6600
AAATTGCACA	CATCGCCTCT	ATGCACTACC	AACATCAAAG	AAGGATCAAA	TATTTGTTTA	6660
ACAAGGACTG	ATAGAGGATG	GTATTGTGAT	AATGCAGGAT	CAGTATCCTT	CTTTCCACAG	6720
GCTGACACTT	GTAAAGTACA	GTCCAATCGA	GTATTTTGTG	ACACTATGAA	CAGTTTGACA	6780
TTACCAAGTG	AAGTCAGCCT	TTGTAACACT	GACATATTCA	ATTCCAAGTA	TGACTGCAAA	6840
ATTATGACAT	CAAAAACAGA	CATAAGCAGC	TCAGTAATTA	CTTCTCTTGG	AGCTATAGTG	6900
TCATGCTATG	GTAAAACTAA	ATGCACTGCA	TCCAACAAAA	ATCGTGGGAT	TATAAAGACA	6960
TTTTCTAATG	GTTGTGACTA	TGTGTCAAAC	aaaggagtag	ATACTGTGTC	AGTGGGCAAC	7020
ACTTTATACT	ATGTAAACAA	GCTGGAAGGC	AAGAACCTTT	ATGTAAAAGG	GGAACCTATA	7080
ATAAATTACT	ATGACCCTCT	AGTGTTTCCT	TCTGATGAGT	TTGATGCATC	AATATCTCAA	7140
GTCAATGAAA	AAATCAATCA	AAGTTTAGCT	TTTATTCGTA	GATCTGATGA	ATTACTACAT	7200
AATGTAAATA	CTGGCAAATC	TACTACAAAT	ATTATGATAA	CTACAATTAT	TATAGTAATC	7260
ATTGTAGTAT	TGTTATCATT	AATAGCTATT	GGTTTACTGT	TGTATTGTAA	AGCCAAAAAC	7320
ACACCAGTTA	CACTAAGCAA	AGACCAACTA	AGTGGAATCA	ATAATATTGC	ATTCAGCAAA	7380

- 297 -

TAGACAAAAA ACCACCTGAT CATGTTTCAA CAACAATCTG CTGACCACCA ATCCCAAATC	7440
AACTTACAAC AAATATTTCA ACATCACAGT ACAGGCTGAA TCATTTCCTC ACATCATGCT	7500
ACCCACATAA CTAAGCTAGA TCCTTAACTT ATAGTTACAT AAAAACCTCA AGTATCACAA	7560
TCAACCACTA AATCAACACA TCATTCACAA AATTAACAGC TGGGGCAAAT ATGTCGCGAA	7620
GAAATCCTTG TAAATTTGAG ATTAGAGGTC ATTGCTTGAA TGGTAGAAGA TGTCACTACA	7680
GTCATAATTA CTTTGAATGG CCTCCTCATG CATTACTAGT GAGGCAAAAC TTCATGTTAA	7740
ACAAGATACT CAAGTCAATG GACAAAAGCA TAGACACTTT GTCTGAAATA AGTGGAGCTG	7800
CTGAACTGGA TAGAACAGAA GAATATGCTC TTGGTATAGT TGGAGTGCTA GAGAGTTACA	7860
TAGGATCTAT AAACAACATA ACAAAACAAT CAGCATGTGT TGCTATGAGT AAACTTCTTA	7920
TTGAGATCAA TAGTGATGAC ATTAAAAAGC TTAGAGATAA TGAAGAACCC AATTCACCTA	7980
AGATAAGAGT GTACAATACT GTTATATCAT ACATTGAGAG CAATAGAAAA AACAACAAGC	8040
AAACCATCCA TCTGCTCAAG AGACTACCAG CAGACGTGCT GAAGAAGACA ATAAAGAACA	8100
CATTAGATAT CCACAAAAGC ATAACCATAA GCAATCCAAA AGAGTCAACT GTGAATGATC	8160
AAAATGACCA AACCAAAAAT AATGATATTA CCGGATAAAT ATCCTTGTAG TATATCATCC	8220
ATATTGATCT CAAGTGAAAG CATGGTTGCT ACATTCAATC ATAAAAACAT ATTACAATTT	8280
AACCATAACT ATTTGGATAA CCACCAGCGT TTATTAAATC ATATATTTGA TGAAATTCAT	8340
TGGACACCTA AAAACTTATT AGATGCCACT CAACAATTTC TCCAACATCT TAACATCCCT	8400
GAAGATATAT ATACAGTATA TATATTAGTG TCATAATGCT TGACCATAAC GACTCTATGT	B460
CATCCAACCA TAAAACTATT TTGATAAGGT TATGGGACAA AATGGATCCC ATTATTAATG	8520
GAAACTCTGC TAATGTGTAT CTAACTGATA GTTATTTAAA AGGTGTTATC TCTTTTTCAG	8580
AGTGTAATGC TTTAGGGAGT TATCTTTTTA ACGGCCCTTA TCTTAAAAAT GATTACACCA	8640
ACTTANTING TAGACAAAGC CCACTACTAG AGCATATGAA TCTTAAAAAA CTAACTATAA	8700
CACAGTCATT AATATCTAGA TATCATAAAG GTGAACTGAA ATTAGAAGAA CCAACTTATT	8760
TCCAGTCATT ACTTATGACA TATAAAAGTA TGTCCTCGTC TGAACAAATT GCTACAACTA	8820
ACTTACTTAA AAAAATAATA CGAAGAGCCA TAGAAATAAG TGATGTAAAG GTGTACGCCA	8880
TCTTGAATAA ACTAGGATTA AAGGAAAAGG ACAGAGTTAA GCCCAACAAT AATTCAGGTG	8940

- 298 -

ATGAAAACTC	AGTACTTACA	ACCATAATTA	AAGATGATAT	ACTTTCGGCT	GTGGAAAACA	9000
ATCAATCATA	TACAAATTCA	GACAAAAGTC	ACTCAGTAAA	TCAAAATATC	ACTATCAAAA	9060
CAACACTCTT	GAAAAAATTG	ATGTGTTCAA	TGCAACATCC	TCCATCATGG	TTAATACACT	9120
GGTTCAATTT	ATATACAAAA	TTAAATAACA	TATTAACACA	ATATCGATCA	AATGAGGTAA	9180
AAAGTCATGG	GTTTATATTA	ATAGATAATC	AAACTTTAAG	TGGTTTTCAG	TTTATTTTAA	9240
ATCAATATGG	TTGTATCGTT	TATCATAAAG	GACTCAAAAA	AATCACAACT	ACTACTTACA	9300
ATCAATTTT	GACATGGAAA	GACATCAGCC	TTAGCAGATT	AAATGTTTGC	TTAATTACTT	9360
GGATAAGTAA	TTGTTTAAAT	ACATTAAACA	AAAGCTTAGG	GCTGAGATGT	GGATTCAATA	9420
ATGTTGTGTT	ATCACAATTA	TTTCTTTATG	GAGATTGTAT	ACTGAAATTA	TTTCATAATG	9480
AAGGCTTCTA	CATAATAAAA	GAAGTAGAGG	GATTTATTAT	GTCTTTAATT	CTARACATAR	9540
CAGAAGAAGA	TCAATTTAGG	AAACGATTTT	ATAATAGCAT	GCTAAATAAC	ATCACAGATG	9600
CAGCTATTAA	GGCTCAAAAG	GACCTACTAT	CAAGAGTATG	TCACACTTTA	TTAGACAAGA	9660
CAGTGTCTGA	TAATATCATA	AATGGTAAAT	GGATAATCCT	ATTAAGTAAA	TTTCTTAAAT	9720
TGATTAAGCT	TGCAGGTGAT	AATAATCTCA	ATAACTTGAG	TGAGCTATAT	TTTCTCTTCA	9780
GAATCTTTGG	ACATCCAATG	GTCGATGAAA	GACAAGCAAT	GGATTCTGTA	AGAATTAACT	9840
GTAATGAAAC	TAAGTTCTAC	TTATTAAGTA	GTCTAAGTAC	ATTAAGAGGT	GCTTTCATTT	9900
ATAGAATCAT	AAAAGGGTTT	GTAAATACCT	ACAACAGATG	GCCCACCTTA	AGGAATGCTA	9960
TTGTCCTACC	TCTAAGATGG	TTAAACTACT	ATAAACTTAA	TACTTATCCA	TCTCTACTTG	10020
AAATCACAGA	AAATGATTTG	ATTATTTTAT	CAGGATTGCG	GTTCTATCGT	GAGTTTCATC	10080
TGCCTAAAAA	AGTGGATCTT	GAAATGATAA	TAAATGACAA	AGCCATTTCA	CCTCCAAAAG	10140
ATCTAATATG	GACTAGTTTT	CCTAGAAATT	ACATGCCATC	ACATATACAA	AATTATATAG	10200
AACATGAAAA	GTTGAAGTTC	TCTGAAAGCG	ACAGATCGAG	AAGAGTACTA	GAGTATTACT	10260
TGAGAGATAA	TAAATTCAAT	GAATGCGATC	TATACAATTG	TGTAGTCAAT	CAAAGCTATC	10320
TCAACAACTC	TAATCACGTG	GTATCACTAA	CTGGTAAAGA	AAGAGAGCTC	AGTGTAGGTA	10380
GAATGTTTGC	TATGCAACCA	GGTATGTTTA	GGCAAATCCA	AATCTTAGCA	GAGAAAATGA	10440
TAGCTGAAAA	TATTTTACAA	TTCTTCCCTG	AGAGTTTGAC	AAGATATGGT	GATCTAGAGC	10500

- 299 -

TTCAAAAGAT ATTAGAATTA AAAGCAGGAA TAAGCAACAA GTCAAATCGT TATAATGATA	10560
ACTACAACAA TTATATCAGT AAATGTTCTA TCATTACAGA TCTTAGCAAA TTCAATCAGG	10620
CATTTAGATA TGAAACATCA TGTATCTGCA GTGATGTATT AGATGAACTG CATGGAGTAC	10680
AATCTCTGTT CTCTTGGTTG CATTTAACAA TACCTCTTGT CACAATAATA TGTACATATA	10740
GACATGCACC TCCTTTCATA AAGGATCATG TTGTTAATCT TAATGAGGTT GATGAACAAA	10800
GTGGATTATA CAGATATCAT ATGGGTGGTA TTGAGGGCTG GTGTCAAAAA CTGTGGACCA	10860
TTGAAGCTAT ATCATTATTA GATCTAATAT CTCTCAAAGG GAAATTCTCT ATCACAGCTC	10920
TGATAAATGG TGATAATCAG TCAATTGATA TAAGCAAACC AGTTAGACTT ATAGAGGGTC	10980
AGACCCATGC ACAAGCAGAT TATTTGTTAG CATTAAATAG CCTTAAATTG TTATATAAAG	11040
AGTATGCAGG TATAGGCCAT AAGCTTAAGG GAACAGAGAC CTATATATCC CGAGATATGC	11100
AGTTCATGAG CAAAACAATC CAGCACAATG GAGTGTACTA TCCAGCCAGT ATCAAAAAAG	11160
TCCTGAGAGT AGGTCCATGG ATAAACACGA TACTTGATGA TTTTAAAGTT AGTTTAGAAT	11220
CTATAGGCAG CTTAACACAG GAGTTAGAAT ACAGAGGAGA AAGCTTATTA TGCAGTTTAA	11280
TATTTAGGAA CATTTGGTTA TACAATCAAA TTGCTTTGCA ACTCCGAAAT CATGCATTAT	11340
GTAACAATAA GCTATATTTA GATATATTGA AAGTATTAAA ACACTTAAAA ACTTTTTTTA	11400
ATCTTGATAG CATTGATATG GCTTTATCAT TGTATATGAA TTTGCCTATG CTGTTTGGTG	11460
GTGGTGATCC TAATTTGTTA TATCGAAGCT TTTATAGGAG AACTCCAGAC TTCCTTACAG	11520
AAGCTATAGT ACATTCAGTG TTTGTGTTGA GCTATTATAC TGGTCACGAT TTACAAGATA	11580
AGCTCCAGGA TCTTCCAGAT GATAGACTGA ACAAATTCTT GACATGTGTC ATCACATTTG	11640
ATAAAAATCC CAATGCCGAG TTTGTAACAT TGATGAGGGA TCCACAGGCT TTAGGGTCTG	11700
AAAGGCAAGC TAAAATTACT AGTGAGATTA ATAGATTAGC AGTAACAGAA GTCTTAAGTA	11760
TAGCCCCAAA CAAAATATTT TCTAAAAGTG CACAACATTA TACTACCACT GAGATTGATC	11820
TAAATGACAT TATGCAAAAT ATAGAACCAA CTTACCCTCA TGGATTAAGA GTTGTTTATG	11880
AAAGTTTACC TTTTTATAAA GCAGAAAAAA TAGTTAATCT TATATCAGGA ACAAAATCCA	11940
TAACTAATAT ACTTGAAAAA ACATCAGCAA TAGATACAAC TGATATTAAT AGGGCTACTG	12000
ATATGATGAG GAAAAATATA ACTTTACTTA TAAGGATACT TCCACTAGAT TGTAACAAAG	12060

- 300 -

ACAAAAGAGA	GTTATTAAGT	TTAGAAAATC	TTAGTATAAC	TGAATTAAGC	aagtatgtaa	12120
GAGAAAGATC	TTGGTCATTA	TCCAATATAG	TAGGAGTAAC	ATCGCCAAGT	ATTATGTTCA	12180
CAATGGACAT	TAAATATACA	ACTAGCACTA	TAGCCAGTGG	ATAATAATA	GAAAAATATA	12240
ATGTTAATAG	TTTAACTCGT	GGTGAAAGAG	GACCCACCAA	GCCATGGGTA	GGCTCATCCA	12300
CGCAGGAGAA	AAAAACAATG	CCAGTGTACA	ACAGACAAGT	TTTAACCAAA	AAGCAAAGAG	12360
ACCARATAGA	TTTATTAGCA	AAATTAGACT	GGGTATATGC	ATCCATAGAC	AACAAAGATG	12420
AATTCATGGA	AGAACTGAGT	ACTGGAACAC	TTGGACTGTC	ATATGAAAA	GCCAAAAAGT	12480
TGTTTCCACA	ATATCTAAGT	GTCAATTATT	TACACCGTTT	AACAGTCAGT	AGTAGACCAT	12540
GTGAATTCCC	TGCATCAATA	CCAGCTTATA	GAACAACAAA	TTATCATTTT	GATACTAGTC	12600
CTATCAATCA	TGTATTAACA	GAAAAGTATG	GAGATGAAGA	TATCGACATT	GTGTTTCAAA	12660
ATTGCATAAG	TTTTGGTCTT	AGCCTGATGT	CGGTTGTGGA	ACAATTCACA	AACATATGTC	12720
CTAATAGAAT	TATTCTCATA	CCGAAGCTGA	ATGAGATACA	TTTGATGAAA	CCTCCTATAT	12780
TTACAGGAGA	TGTTGATATC	ATCAAGTTGA	AGCAAGTGAT	ACAAAAGCAG	CACATGTTCC	12840
TACCAGATAA	AATAAGTTTA	ACCCAATATG	TAGAATTATT	CTTAAGTAAC	AAAGCACTTA	12900
AATCTGGATC	TCACATCAAC	TCTAATTTAA	TATTAGTACA	TAAAATGTCT	GATTATTTTC	12960
ATAATGCTTA	TATTTTAAGT	ACTAATTTAG	CTGGACATTG	GATTCTGATT	ATTCAACTTA	13020
TGAAAGATTC	AAAAGGTATT	TTTGAAAAAG	ATTGGGGAGA	GGGGTACATA	ACTGATCATA	13080
TGTTCATTAA	TTTGAATGTT	TTCTTTAATG	CTTATAAGAC	TTATTTGCTA	TGTTTTCATA	13140
AAGGTTATGG	TAAAGCAAAA	TTAGAATGTG	ATATGAACAC	TTCAGATCTT	CTTTGTGT <b>T</b> T	13200
TGGAGTTAAT	AGACAGTAGC	TACTGGAAAT	CTATGTCTAA	AGTTTTCCTA	GAACAAAAAG	13260
TCATAAAATA	CATAGTCAAT	CAAGACACAA	GTTTGCGTAG	AATAAAAGGC	TGTCACAGTT	13320
TTAAGTTGTG	GTTTTTAAAA	CGCCTTAATA	ATGCTAAATT	TACCGTATGC	CCTTGGGTTG	13380
TTAACATAGA	TTATCACCCA	ACACACATGA	AAGCTATATT	ATCTTACATA	GATTTAGTTA	13440
GAATGGGGTT	AATAAATGTA	GATAAATTAA	CCATTAAAAA	TAAAAACAAA	TTCAATGATG	13500
AATTTTACAC	ATCAAATCTC	TTTTACATTA	GTTATAACTT	TTCAGACAAC	ACTCATTTGC	13560
TAACAAAACA	AATAAGAATT	GCTAATTCAG	AATTAGAAGA	TAATTATAAC	AAACTATATC	13620

- 301 -

ACCCAACCCC AGAAACTTTA GAAAATATGT CATTAATTCC TGTTAAAAGT AATAATAGTA	13680
ACAAACCTAA ATTTTGTATA AGTGGAAATA CCGAATCTAT GATGATGTCA ACATTCTCTA	13740
GTANAATGCA TATTAAATCT TCCACTGTTA CCACAAGATT CAATTATAGC AAACAAGACT	13800
TGTACAATTT ATTTCCAATT GTTGTGATAG ACAAGATTAT AGATCATTCA GGTAATACAG	13860
CAARATCTAA CCAACTTTAC ACCACCACTT CACATCAGAC ATCTTTAGTA AGGAATAGTG	13920
CATCACTITA TIGCATGCIT CCTTGGCATC ATGTCAATAG ATTTAACTIT GTATTTAGTT	13980
CCACAGGATG CAAGATCAGT ATAGAGTATA TTTTAAAAGA TCTTAAGATT AAGGACCCCA	14040
GTTGTATAGC ATTCATAGGT GAAGGAGCTG GTAACTTATT ATTACGTACG GTAGTAGAAC	14100
TTCATCCAGA CATAAGATAC ATTTACAGAA GTTTAAAAGA TTGCAATGAT CATAGTTTAC	14160
CTATTGAATT TCTAAGGTTA TACAACGGGC ATATAAACAT AGATTATGGT GAGAATTTAA	14220
CCATTCCTGC TACAGATGCA ACTAATAACA TTCATTGGTC TTATTTACAT ATAAAATTTG	14280
CAGAACCTAT TAGCATCTTT GTCTGCGATG CTGAATTACC TGTTACAGCC AATTGGAGTA	14340
AAATTATAAT TGAATGGAGT AAGCATGTAA GAAAGTGCAA GTACTGTTCT TCTGTAAATA	14400
GATGCATTTT AATTGCAAAA TATCATGCTC AAGATGACAT TGATTTCAAA TTAGATAACA	14460
TTACTATATT AAAAACTTAC GTGTGCCTAG GTAGCAAGTT AAAAGGATCT GAAGTTTACT	14520
TAATCCTTAC AATAGGCCCT GCAAATATAC TTCCTGTTTT TGATGTTGTA CAAAATGCTA	14580
AATTGACACT TTCAAGAACT AAAAATTTCA TTATGCCTAA AAAAACTGAC AAGGAATCTA	14640
TCGATGCAAA TATTAAAAGC TTAATACCTT TCCTTTGTTA CCCTATAACA AAAAAAGGAA	14700
TTAAGACTTC ATTGTCAAAA TTGAAGAGTG TAGTTAATGG AGATATATTA TCATATTCTA	14760
TAGCTGGACG TAATGAAGTA TTCAGCAACA AGCTTATAAA CCACAAGCAT ATGAATATCC	14820
TAAAATGGCT AGATCATGTT TTAAATTTTA GATCAGCTGA ACTTAATTAC AATCATTTAT	14880
ACATGATAGA GTCCACATAT CCTTACTTAA GTGAATTGTT AAATAGTTTA ACAACCAATG	14940
AGCTCAAGAA GCTGATTAAA ATAACAGGTA GTGTGCTATA CAACCTTCCC AACGAACAGT	15000
AGTITAAAAT ATCATTAACA AGTITGGTCA AATTTAGATG CTAACACATC ATTATATTAT	15060
AGTTATTAAA AAATATACAA ACTTTTCAAT AATTTAGCAT ATTGATTCCA AAATTATCAT	15120
TTTAGTCTTA AGGGGTTAAA TAAAAGTCTA AAACTAACAA TTATACATGT GCATTCACAA	15180

- 302 -

## CACAACGAGA CATTAGTTTT TGACACTTTT TTTCTCGT

15218

- (2) INFORMATION FOR SEQ ID NO:24:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2166 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15

Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly 20 25 30

Ser Tyr Leu Phe Asn Gly Pro Tyr Leu Lys Asn Asp Tyr Thr Asn Leu 35 40

Ile Ser Arg Gln Ser Pro Leu Leu Glu His Met Asn Leu Lys Lys Leu 50 60

Thr Ile Thr Gln Ser Leu Ile Ser Arg Tyr His Lys Gly Glu Leu Lys 65 70 75 80

Leu Glu Glu Pro Thr Tyr Phe Gln Ser Leu Leu Met Thr Tyr Lys Ser 85 90 95

Met Ser Ser Glu Gln Ile Ala Thr Thr Asn Leu Leu Lys Lys Ile 100 105 110

Ile Arg Arg Ala Ile Glu Ile Ser Asp Val Lys Val Tyr Ala Ile Leu 115 120 125

Asn Lys Leu Gly Leu Lys Glu Lys Asp Arg Val Lys Pro Asn Asn Asn 130 135 140

Ser Gly Asp Glu Asn Ser Val Leu Thr Thr Ile Ile Lys Asp Asp Ile 145 150 155 160

Leu Ser Ala Val Glu Asn Asn Gln Ser Tyr Thr Asn Ser Asp Lys Ser 165 170 175

His Ser Val Asn Gln Asn Ile Thr Ile Lys Thr Thr Leu Leu Lys Lys

- 303 -

			180	)				185	5				190	)	
Leu	Met	Cys 195		Met	: Gln	His	200		Sez	Tr	) Let	1 Ile 205		Tr	Ph.
Asn	Leu 210		Thr	Lys	Leu	215		Ile	Leu	ı Thr	Glr 220		Arg	, Ser	aA :
Glu 225		Lys	Ser	His	Gly 230		Ile	Leu	Ile	235		Glr	Thr	Lev	24
Gly	Phe	Gln	Phe	11e 245	Leu	Asn	Gln	Tyr	Gly 250		: Ile	Va]	Туг	His 255	-
Gly	Leu	Lys	Lys 260		Thr	Thr	Thr	Thr 265		Asn	Gln	Phe	270		Tr
Lys	Ąsp	Ile 275		Leu	Ser	Arg	Leu 280	Asn	Val	Сув	Leu	11e 285		Trp	Ile
Ser	Asn 290	Сув	Leu	Asn	Thr	Leu 295	Asn	Lys	Ser	Leu	Gly 300		Arg	Сув	Gly
Phe 305	Asn	Asn	Val	Val	Leu 310	Ser	Gln	Leu	Phe	Leu 315		Gly	Asp	Cys	11e 320
Leu	Lys	Leu	Phe	His 325	Asn	Glu	Gly	Phe	Tyr 330	Ile	Ile	Lys	Glu	Val 335	Glu
Gly	Phe	Ile	Met 340	Ser	Leu	Ile	Leu	Asn 345	Ile	Thr	Glu	Glu	<b>Asp</b> 350	Gln	Phe
Arg	Lys	Arg 355	Phe	Tyr	Asn	Ser	Met 360	Leu	Asn	Asn	Ile	Thr 365	Asp	Ala	Ala
	370				Asp	375					380				
385					<b>Asp</b> 390					395					400
Leu	Ser	Lys	Phe	Leu 405	Lys	Leu	Ile	Lys	Leu 410	Ala	Gly	Asp	Asn	Asn 415	Leu
Asn	asa	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
		435			Gln		440					445			
	Thr 450	Lys	Phe	Tyr	Leu	Leu 455	Ser	Ser	Leu	Ser	Thr 460	Leu	Arg	Gly	Ala

PCT/US97/16718 WO 98/13501

- 304 -

Phe 465	Ile	Туг	Arg	Ile	Ile 470	Lys	Gly	Phe	Val	Asn 475	Thr	Tyr	Asn	Arg	Trp 480
Pro	Thr	Leu	Arg	Asn 485	Ala	Ile	Val	Leu	Pro 490	Leu	Arg	Trp	Leu	Asn 495	Tyr
Tyr	Lys	Leu	Asn 500	Thr	Tyr	Pro	Ser	Leu 505	Leu	Glu	Ile	Thr	Glu 510	Asn	Asp
Leu	Ile	Ile 515	Leu	Ser	Gly	Leu	Arg 520	Phe	Tyr	Arg	Glu	Phe 525	His	Leu	Pro
_	530		Явр			535					540				
545			Leu		550					555					560
			Asn	565					570					575	
_			Arg 580					585					590		
		595	Asp				600					605			
	610		His			615					620				
625			Met		630					635					640
			Glu	645					650					655	
			660					665					670		Glu
		675					680					685			Tyr
	690		Ile			695					700				
705					710					715					120
Asp	Glu	Leu	His	Gly 725		Gln	Ser	Leu	730		Trp	Leu	His	Leu 735	Thr

- 305 -

Ile	Pro	Let			: Ile	ıle	в Суа			r Arg	Hi:	a Ala			o Ph
			740					745					750		
Ile	Lys	755		Va]	Va]	l Ası	760		ı Glu	ı Val	. As <u>ı</u>	76!		ı Se	r Gly
Leu	770		ј Туг	His	Met	: Gly 775		/ Ile	Glu	ı Gly	780		Glr.	Ly:	s Le
Trp 785		Ile	Glu	Ala	790		Lev	ı Lev	a Asr	795		Ser	. Lev	Lys	3 Gl <sub>3</sub> 800
Lys	Phe	Ser	Ile	Thr 805		Lev	lle	) Asn	61y 810	/ Asp	Asn	Glr	ser	11e 815	
Ile	Ser	Lys	820		Arg	Leu	Ile	61u 825		Gln	Thr	His	830		Ala
Asp	Tyr	Leu 835		Ala	Leu	Asn	Ser 840		Lys	Leu	Leu	Tyr 845		Glu	Туг
Ala	Gly 850		Gly	His	Lys	Leu 855		Gly	Thr	Glu	Thr 860		Ile	Ser	Arg
Авр 865		Gln	Phe	Met	Ser 870	Lys	Thr	Ile	Gln	His 875	Asn	Gly	Val	Tyr	Tyr 880
Pro	Ala	Ser	Ile	<b>Lys</b> 885	Lys	Val	Leu	Arg	Val 890	Gly	Pro	Trp	Ile	Asn 895	
Ile	Leu	Asp	<b>qa</b> 900	Phe	Lys	Val	Ser	Leu 905	Glu	Ser	Ile	Gly	Ser 910	Leu	Thr
Gln	Glu	Leu 915	Glu	Tyr	Arg	Gly	Glu 920	Ser	Leu	Leu	Сув	Ser 925	Leu	Ile	Phe
Arg	<b>Asn</b> 930	Ile	Trp	Leu	Tyr	Asn 935	Gln	Ile	Ala	Leu	Gln 940	Leu	Arg	Asn	His
Ala 945	Leu	Сув	Asn		<b>Lув</b> 950		Tyr	Leu		Ile 955		Lys	Val	Leu	<b>Lys</b> 960
His	Leu	Lys	Thr	Phe 965	Phe	Asn	Leu	Asp	Ser 970	Ile	Asp	Met	Ala	Leu 975	Ser
Leu	Tyr	Met	Asn 980	Leu	Pro	Met	Leu	Phe 985	Gly	Gly	Gly	Asp	Pro 990	Asn	Leu
Leu	Tyr	Arg 995	Ser	Phe	Tyr	Arg	Arg 1000		Pro	Asp	Phe	Leu 1005		Glu	Ala
Ile	Val	His	Ser	Val	Phe	Val	Leu	Ser	Tyr	Tyr	Thr	Gly	His	Asp	Leu

- 306 -

	1010	,				1013	•				1020	,			
Gln 1025	Asp	Lys	Leu	Gln	Asp 1030		Pro	Asp	Asp	Arg 1035		Asn	Lys	Phe	Leu 1040
Thr	Сув	Val	Ile	Thr 1045		Asp	Lys	Asn	Pro 1050		Ala	Glu	Phe	Val 1055	
Leu	Met	Arg	Asp 1060		Gln	Ala	Leu	Gly 1065		Glu	Arg	Gln	Ala 1070		Ile
Thr	Ser	Glu 1075		Asn	Arg	Leu	Ala 1080		Thr	Glu	Val	Leu 108!		Ile	Ala
Pro	Asn 1090	-	Ile	Phe	Ser	Lув 1095		Ala	Gln	His	Tyr 110		Thr	Thr	Glu
Ile 110	Asp 5	Leu	Asn	Авр	1110		Gln	Asn	Ile	Glu 1115		Thr	Tyr	Pro	His 1120
Gly	Leu	Arg	Val	Val 112	_	Glu	Ser	Leu	Pro 1130		Tyr	Lys	Ala	Glu 1135	
Ile	Val	Asn	Leu 1140		Ser	Gly	Thr	Lys 114		Ile	Thr	Asn	Ile 1150		Glu
Lys	Thr	Ser 115		Ile	Asp	Thr	Thr 116		Ile	Asn	Arg	Ala 116		Asp	Met
Met	Arg 117	_	Asn	Ile	Thr	Leu 117		Ile	Arg	Ile	Leu 118		Leu	Asp	Сув
Asn 118	Lys 5	Дар	Lys	Arg	Glu 119		Leu	Ser	Leu	Glu 119		Leu	Ser	Ile	Thr 120
Glu	Leu	Ser	Lys	Tyr 120		Arg	Glu	Arg	Ser 121		Ser	Leu	Ser	Asn 121	
Val	Gly	Val	Thr 122		Pro	Ser	Ile	Met 122		Thr	Met	Asp	11e 123		Tyr
Thr	Thr	Ser 123		Ile	Ala	Ser	Gly 124		Ile	Ile	Glu	Lys 124		Asn	Val
Asn	Ser 125		Thr	Arg	Gly	Glu 125		Gly	Pro	Thr	Lув 126		Trp	Val	Gly
Ser 126	Ser 5	Thr	Gln	Glu	Lys 127		Thr	Met	Pro	Val 127		Asn	Arg	Gln	Val 128
Leu	Thr	Lys	Lys	Gln 128		Asp	Gln		Asp		Leu	Ala	Lys	Leu 129	

- 307 -

- Trp Val Tyr Ala Ser Ile Asp Asn Lys Asp Glu Phe Met Glu Glu Leu 1300 1305 1310
- Ser Thr Gly Thr Leu Gly Leu Ser Tyr Glu Lys Ala Lys Lys Leu Phe 1315 1320 1325
- Pro Gln Tyr Leu Ser Val Asn Tyr Leu His Arg Leu Thr Val Ser Ser 1330 1335 1340
- Arg Pro Cys Glu Phe Pro Ala Ser Ile Pro Ala Tyr Arg Thr Thr Asn 1345 1350 1355 1360
- Tyr His Phe Asp Thr Ser Pro Ile Asn His Val Leu Thr Glu Lys Tyr 1365 1370 1375
- Gly Asp Glu Asp Ile Asp Ile Val Phe Gln Asn Cys Ile Ser Phe Gly 1380 1385 1390
- Leu Ser Leu Met Ser Val Val Glu Gln Phe Thr Asn Ile Cys Pro Asn 1395 1400 1405
- Arg Ile Ile Leu Ile Pro Lys Leu Asn Glu Ile His Leu Met Lys Pro 1410 1415 1420
- Pro Ile Phe Thr Gly Asp Val Asp Ile Ile Lys Leu Lys Gln Val Ile 1425 1430 1435 1440
- Gln Lys Gln His Met Phe Leu Pro Asp Lys Ile Ser Leu Thr Gln Tyr 1445 1450 1455
- Val Glu Leu Phe Leu Ser Asn Lys Ala Leu Lys Ser Gly Ser His Ile 1460 1465 1470
- Asn Ser Asn Leu Ile Leu Val His Lys Met Ser Asp Tyr Phe His Asn 1475 1480 1485
- Ala Tyr Ile Leu Ser Thr Asn Leu Ala Gly His Trp Ile Leu Ile Ile 1490 1495 1500
- Gln Leu Met Lys Asp Ser Lys Gly Ile Phe Glu Lys Asp Trp Gly Glu 1505 1510 1515 1520
- Gly Tyr Ile Thr Asp His Met Phe Ile Asn Leu Asn Val Phe Phe Asn 1525 1530 1535
- Ala Tyr Lys Thr Tyr Leu Leu Cys Phe His Lys Gly Tyr Gly Lys Ala 1540 1545 1550
- Lys Leu Glu Cys Asp Met Asn Thr Ser Asp Leu Cys Val Leu Glu 1555 1560 1565

- 308 -

- Leu Ile Asp Ser Ser Tyr Trp Lys Ser Met Ser Lys Val Phe Leu Glu 1570 1575 1580
- Gln Lys Val Ile Lys Tyr Ile Val Asn Gln Asp Thr Ser Leu Arg Arg 1585 1590 1595 1600
- Ile Lys Gly Cys His Ser Phe Lys Leu Trp Phe Leu Lys Arg Leu Asn 1605 1610 1615
- Asn Ala Lys Phe Thr Val Cys Pro Trp Val Val Asn Ile Asp Tyr His 1620 1625 1630
- Pro Thr His Met Lys Ala Ile Leu Ser Tyr Ile Asp Leu Val Arg Met 1635 1640 1645
- Gly Leu Ile Asn Val Asp Lys Leu Thr Ile Lys Asn Lys Asn Lys Phe 1650 1660
- Asn Asp Glu Phe Tyr Thr Ser Asn Leu Phe Tyr Ile Ser Tyr Asn Phe 1665 1670 1675 1680
- Ser Asp Asn Thr His Leu Leu Thr Lys Gln Ile Arg Ile Ala Asn Ser 1685 1690 1695
- Glu Leu Glu Asp Asn Tyr Asn Lys Leu Tyr His Pro Thr Pro Glu Thr
- Leu Glu Asn Met Ser Leu Ile Pro Val Lys Ser Asn Asn Ser Asn Lys 1715 1720 1725
- Pro Lys Phe Cys Ile Ser Gly Asn Thr Glu Ser Met Met Ser Thr 1730 1735 1740
- Phe Ser Ser Lys Met His Ile Lys Ser Ser Thr Val Thr Thr Arg Phe 1745 1750 1755 1760
- Asn Tyr Ser Lys Gln Asp Leu Tyr Asn Leu Phe Pro Ile Val Val Ile 1765 1770 1775
- Asp Lys Ile Ile Asp His Ser Gly Asn Thr Ala Lys Ser Asn Gln Leu 1780 1785 1790
- Tyr Thr Thr Thr Ser His Gln Thr Ser Leu Val Arg Asn Ser Ala Ser 1795 1800 1805
- Leu Tyr Cys Met Leu Pro Trp His His Val Asn Arg Phe Asn Phe Val 1810 1815 1820
- Phe Ser Ser Thr Gly Cys Lys Ile Ser Ile Glu Tyr Ile Leu Lys Asp 1825 1830 1835 1840
- Leu Lys Ile Lys Asp Pro Ser Cys Ile Ala Phe Ile Gly Glu Gly Ala

- 309 -

	1845	1850		1855
Gly Asn Leu Le 18	eu Leu Arg Th	r Val Val Glu Leu 1865	His Pro Asp 1870	
Tyr Ile Tyr Az 1875	g Ser Leu Ly	s Asp Cys Asn Asp 1880	His Ser Leu 1885	Pro Ile
Glu Phe Leu Ar 1890	g Leu Tyr Ası 189	n Gly His Ile Asn 95	Ile Asp Tyr (	Gly Glu
1905	1910	r Asp Ala Thr Asn 191	5	1920
Tyr Leu His Il	e Lys Phe Ala 1925	a Glu Pro Ile Ser 1930		Сув Авр 1935
Ala Glu Leu Pr 19	o Val Thr Ala 40	Asn Trp Ser Lys 1945	Ile Ile Ile G 1950	Slu Trp
Ser Lys His Va 1955	l Arg Lys Cys	Lys Tyr Cys Ser 1960	Ser Val Asn A 1965	rg Cys
Ile Leu Ile Ala 1970	a Lys Tyr His 197	Ala Gln Asp Asp 5	Ile Asp Phe L 1980	ys Leu
1965	1990	Thr Tyr Val Cys 1995		2000
	2005	Ile Leu Thr Ile 2010	20	015
202	:0	Gln Asn Ala Lys : 2025	2030	
2035		Lys Lys Thr Asp 1 2040	2045	_
7 2030	2055		3060	
2005	2070	Ser Lys Leu Lys S 2075		2080
	2085	Ala Gly Arg Asn 6 2090	20	95
2100	)	Met Asn Ile Leu L 2105	2110	
Val Leu Asn Phe	Arg Ser Ala	Glu Leu Asn Tyr A	sn His Leu Ty	r Met

2120

2115

- 310 -

Ile Glu Ser Thr Tyr Pro Tyr Leu Ser Glu Leu Leu Asn Ser Leu Thr 2130 2135 2140

Thr Asn Glu Leu Lys Lys Leu Ile Lys Ile Thr Gly Ser Val Leu Tyr 2145 2150 2155 2160

Asn Leu Pro Asn Glu Gln 2165

#### (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15229 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA (genomic)

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACGCGAAAAA ATGCGTACTA CAAACTTGCA CATTCGGAAA AAATGGGGCA AATAAGAATT 60 TGATAAGTGC TATTTAAATC TAACCTTTTC AATCAGAAAT GGGGTGCAAT TCACTGAGCA 120 TGATAAAGGT TAGATTACAA AATTTATTTG ACAATGACGA AGTAGCATTG TTAAAAATAA 180 CATGTTATAC TGACAAATTA ATTCTTCTGA CCAATGCATT AGCCAAAGCA GTAATACATA 240 CARTTARATT ARACGGCATA GTTTTTATAC ATGTTATAAC ARGCAGTGAA GTGTGCCCTG 300 360 ACAACAATAT TGTAGTGAAA TCTAACTTTA CAACAATGCC AATATTACAA AACGGAGGAT ACATATGGGA ATTGATTGAG TTGACACACT GCTCTCAATC AAATGGTCTA ATGGATGATA 420 ATTGTGAAAT CAAATTTTCT AAAAGACTAA GTGACTCAGT AATGACTAAT TATATGAATC 480 540 AAATATCTGA TTTACTTGGG CTTGATCTCA ATTCATGAAT TATGTTTAGT CTAATTTAAT AGACATGTGT TTATCACCAT TTTAGTTAAT ATAAAACCTC ATCAAAGGGA AATGGGGCAA 600 ATANACTCAC CTAATCAGTC ANACCATGAG CACTACANAT GACAACACTA CTATGCANAG 660 ATTGATGATC ACAGACATGA GACCCCTGTC GATGGAATCA ATAATAACAT CTCTCACCAA 720 AGAAATCATA ACACACAAAT TCATATACTT GATAAACAAT GAATGTATTG TAAGAAAACT 780 TGATGAAAGA CAAGCTACAT TTACATTCTT AGTCAATTAT GAGATGAAGC TATTGCACAA 840

- 311 -

AGTAGGGAGT	ACCAAATACA	AGAAATACAC	: TGAATATAAT	ACAAAATATG	GCACTTTCCC	900
CATGCCTATA	TTTATCAATC	ATGACGGGTT	TCTAGAATGT	ATTGGCATTA	AGCCTACAAA	960
ACACACTCCT	ATAATATACA	AATATGACCT	CAACCCGTAA	ATTCCAACAA	AAAACTAACC	1020
CATCCAAACT	AAGCTATTCC	TCAAACAACA	GTGCTCAACA	GTTAAGAAGG	AGCTAATCCA	1080
TTTTAGTAAT	TAAAAATAAA	GGCAGAGCCA	ATAACATAAA	TTGGGGCAAA	TACAAAGATG	1140
GCTCTTAGCA	AAGTCAAGTT	AAATGATACA	TTAAATAAGG	ATCAGCTGCT	GTCATCCAGC	1200
AAATACACTA	TTCAACGTAG	TACAGGAGAT	AATATTGAGA	CTCCCAATTA	TGATGTGCAA	1260
AAACACCTAA	ACAAACTATG	TGGTATGCTA	TTAATCACTG	AAGATGCAAA	TCATAAATTC	1320
ACAGGATTAA	TAGGTATGTT	ATATGCTATG	TCCAGGTTAG	GAAGGGAAGA	CACTATAAAG	1380
ATACTTAAAG	ATGCTGGATA	TCATGTTAAA	GCTAATGGAG	TAGATATAAC	AACATATCGT	1440
CAAGATATAA	ACGGAAAGGA	AATGAAATTC	GAAGTATTAA	CATTATCAAG	CTTGACATCA	1500
GAAATACAAG	TCAATATTGA	GATAGAATCT	AGAAAGTCCT	ACAAAAAAT	GCTAAAAGAG	1560
ATGGGAGAAG	TGGCTCCAGA	ATATAGGCAT	GATTCTCCAG	ACTGTGGGAT	GATAATACTG	1620
TGTATAGCTG	CACTTGTAAT	AACCAAGTTA	GCAGCAGGAG	ATAGATCAGG	TCTTACAGCA	1680
GTAATTAGGA	GGGCAAACAA	TGTCTTAAAA	AACGAAATAA	AACGCTACAA	GGGCCTCATA	1740
CCAAAGGATA	TAGCTAACAG	TTTTTATGAA	GTGTTTGAAA	AACACCCTCA	TCTTATAGAT	1800
GTTTTTGTGC	ACTTTGGCAT	TGCACAATCA	TCCACAAGAG	GGGGTAGTAG	agttgaagga	1860
ATCTTTGCAG	GATTATTTAT	GAATGCCTAT	GGTTCAGGGC	AAGTAATGCT	AAGATGGGGA	1920
GTTCTAGCCA	AATCTGTAAA	AAATATCATG	CTAGGACATG	CTAGTGTCCA	GGCAGAAATG	1980
GAACAAGTTG	TGGAAGTTTA	TGAGTATGCA	CAGAAGTTGG	GAGGAGAAGC	TGGATTCTAC	2040
CATATATTGA	ACAATCCAAA	AGCATCATTG	CTGTCATTAA	CTCAATTTCC	TAACTTCTCA	2100
AGTGTGGTCC	TAGGCAATGC	AGCAGGTCTA	GGCATAATGG	GAGAGTATAG	AGGTACACCA	2160
AGAAACCAAG	ATCTATATGA	TGCAGCCAAA	GCATATGCAG	AGCAACTCAA	AGAAAATGGA	2220
GTAATAAACT	ACAGTGTATT	AGACTTAACA	GCAGAAGAAT	TGGAAGCCAT	AAAGCATCAA	2280
CTCAACCCCA	AAGAAGATGA	TGTAGAGCTT	TAAGTTAACA	AAAAATACGG	GGCAAATAAG	2340
TCAACATGGA	GAAGTTTGCA	CCTGAATTTC	atggagaaga	TGCAAACAAC .	AAAGCTACCA	2400

- 312 -

AATTCCTAGA	ATCAATAAAG	GGCAAGTTTG	CATCATCCAA	AGATCCTAAG	AAGAAAGATA	2460
GCATAATATC	TGTTAACTCA	ATAGATATAG	AAGTAACTAA	AGAGAGCCCG	ATAACATCTG	2520
GCACCAACAT	CATCAATCCA	ATAAGTGAAG	CTGATAGTAC	CCCAGAAGCT	AAAGCCAACT	2580
ACCCAAGAAA	ACCCCTAGTA	AGCTTCAAAG	AAGATCTCAC	CCCAAGTGAC	AACCCCTTTT	2640
CTAAGTTGTA	CAAAGAAACA	ATAGAAACAT	TTGATAACAA	TGAAGAAGAA	TCTAGCTACT	2700
CATATGAAGA	AATAAATGAT	CAAACAAATG	ACAACATTAC	AGCAAGACTA	GATAGAATTG	2760
ATGAAAAATT	AAGTGAAATA	TTAGGAATGC	TCCATACATT	AGTAGTTGCA	AGTGCAGGAC	2820
CCACCTCAGC	TCGCGATGGA	ATAAGAGATG	CTATGGTTGG	TCTAAGAGAA	GAAATGATAG	2880
AAAAAATAAG	AGCGGAAGCA	TTAATGACCA	ATGATAGGTT	AGAGGCTATG	GCAAGACTTA	2940
GGAATGAGGA	AAGCGAAAAA	ATGGCAAAAG	ACACCTCAGA	TGAAGTGTCT	CTTAATCCAA	3000
CTTCCAAAAA	ATTGAGTAAT	TTGTTGGAAG	ACAACGATAG	TGACAATGAT	CTATCACTTG	3060
ATGATTTTTG	ATCAGTGATC	AACTCACTCA	GCAATCAACA	ACATCAATGA	AACAGACATC	3120
AATCCATTGA	ATCAACTGCC	AGACTGAACA	CACAAACGTC	CATCAGCAGA	ACTACCAACC	3180
AATCAATCAA	CCAATTGATC	AATCAGCGAC	CTAACAAAAT	TAACAATATA	GTAACAAAAA	3240
AAGAACAAGA	TGGGGCAAAT	ATGGAAACAT	ACGTGAACAA	GCTTCACGAG	GGCTCCACAT	3300
ACACAGCAGC	TGTTCAGTAC	AATGTTCTAG	AAAAAGATGA	TGATCCTGCA	TCACTAACAA	3360
TATGGGTGCC	TATGTTCCAG	TCATCTGTGC	CAGCAGACTT	GCTCATAAAA	GAACTTGCAA	3420
GCATCAACAT	ACTAGTGAAG	CAGATCTCCA	CGCCCAAAGG	ACCTTCACTA	CGAGTCACGA	3480
TTAACTCAAG	AAGTGCTGTG	CTGGCACAAA	TGCCTAGTAG	TTTTATCATA	AGTGCAAATG	3540
TATCATTAGA	TGAAAGAAGC	AAATTAGCAT	ATGATGTAAC	TACACCTTGT	GAAATCAAAG	3600
CATGCAGTCT	AACATGCTTA	AAAGTAAAAA	GTATGTTAAC	TACAGTCAAA	GATCTTACCA	3660
TGAAAACATT	CAATCCCACT	CATGAGATTA	TTGCTCTATG	TGAATTTGAA	AATATTATGA	3720
CATCAAAAAG	AGTAATAATA	CCAACCTATC	TAAGATCAAT	TAGTGTCAAA	AACAAGGACC	3780
TGAACTCACT	AGAAAATATA	GCAACCACCG	AATTCAAAAA	TGCTATCACC	AATGCGAAAA	3840
TTATTCCCTA	TGCAGGATTA	GTATTAGTTA	TCACAGTTAC	TGACAATAAA	GGAGCATTCA	3900
AATATATCAA	GCCACAGAGT	CAATTTATAG	TAGATCTTGG	GGCCTACCTA	GAAAAAGAGA	3960

- 313 -

GCATATATTA TGTGACTACA AATTGGAAGC ATACAGCTAC ACGTTTTTCA ATCAAACC	AC 4020
TAGAGGATTA AACTTAATTA TCAACACTAA ATGACAGGTC CACATATATC TTCAAACT	'AT 4080
ACATTATATC CAAACATCAT GAGCATTTAC ACTACACACT TTTACCATAT AAATCAAT	CT 4140
CATTTAAAAT CCAAAATTAC TTCCAGCTAT CATCTGTTAG ACCTAGAGTG CGAATAGG	TA 4200
AATAAAACCA AAATATGGGG TAAATAGACA TTAGTTAGAG TTCAATCAAT CTCAACAA	CC 4260
ATTTATACCG CCAATTCAGT ACATATACTA TAAATCTCAA AATGGGAAAT ACATCCAT	CA 4320
CARTAGAATT CACAAGCAAA TTTTGGCCTT ATTTTACACT AATACATATG ATCTTAAC	TC 4380
TAATCTCTTT ACTAATTATA ATCACTATTA TGATTGCAAT ACTAAATAAG CTAAGTGA	AC 4440
ATAAAACATT CTGCAACAAA ACTCTTGAAC TAGGACAGAT GTATCAAATC AACACATAG	<b>3</b> T 4500
GTTCTACCAT TATGCTGTGT CAAATTATAA TCTTGTATAT ATAAACAAAC AAATCCAA	rc 4560
TTCTCACAGA GTCATGGTGG CGCAAAACCA CGCCAACCAT CATGATAGCA TAGAGTAGT	TT 4620
ATTTAAAAAT TAACATAATG ATGAATTATT GGTATGAGAT CAGGAACAAC ATTGGGGCA	A 4680
ATGCAGCCAT GTCCAAGCAC AAGAATCGGC GCACTGCCGG GACTCTAGAA AGGACCTGG	G 4740
ATACTCTTAA TCATCTAATT GTAATATCCT CTTGTTTATA CAGATTAAAT TTAAAATCI	A 4800
TAGCACAAAT AGCACTGTCA GTTTTGGCAA TGATAATCTC AACCTCTCTC ATAATTGCA	.G 4860
CCATAATATT CATCATCTCT GCCAATCACA AAGTTACACT AACAACGGTT ACAGTTCAA	A 4920
CAATAAAAA CCACACTGAA AAAAACATCT CCACCTACCT TACTCAAGTC CCACCAGAA	A 4980
GGGTCAACTC ATCCAAACAA CCCACAACCA CATCACCAAT CCACACAAAT TCAGCCACA	A 5040
TATCACCAAA TACAAAATCA GAAACACACC ATACAACAGC ACAAACCAAA GGCAGAATC	A 5100
CCACTTCAAC ACAGACCAAC AAGCCAAGCA CAAAATCACG TTCAAAAAAT CCACCAAAA	A 5160
AACCAAAAGA TGATTACCAT TTTGAAGTGT TCAATTTTGT TCCCTGTAGT ATATGTGGT	A 5220
ATANTCAACT CTGCAAATCC ATCTGCAAAA CAATACCAAG CAACAAACCA AAGAAAAAA	5280
CAACCATCAA ACCCACAAC AAACCAACCA CCAAAACCAC AAACAAAAGA GACCCCAAAI	A 5340
CACCAGCCAA AATGCCAAAA AAAGAAATCA TCACCAACCC AGCAAAAAAA CCAACCCTC	A 5400
AGACCACAGA AAGAGACACC AGCATTTCAC AATCCACCGT GCTCGACACA ATCACTCCAA	5460
AATACACAAT CCAACAGCAA TCCCTCCACT CAACCACCTC CGAAAACACA CCCAGCTCCA	5520

- 314 -

CACAAATACC	CACAGCATCC	GAGCCCTCCA	CATTAAATCC	AAAATTAAAA	ACCTAGTCAC	5580
ATGCTTAGTT	ATTCAAAAAC	TACATCTTAG	CAGAGAACCG	TGATCTATCA	AGCAAGAACA	5640
AAATTAAACC	TGGGGCAAAT	AACCATGGAG	TTGCTGATCC	ACAGGTCAAG	TGCAATCTTC	5700
CTAACTCTTG	CTGTTAATGC	ATTGTACCTC	ACCTCAAGTC	AGAACATAAC	TGAGGAGTTT	5760
TACCAATCGA	CATGTAGTGC	AGTTAGCAGA	GGTTATTTTA	GTGCTTTAAG	AACAGGTTGG	5820
TATACCAGTG	TCATAACAAT	AGAATTAAGT	AATATAAAAG	AAACCAAATG	CAATGGAACT	5880
GACACTAAAG	TAAAACTTAT	AAAACAAGAA	TTAGATAAGT	ATAAGAATGC	AGTAACAGAA	5940
TTACAGCTAC	TTATGCAAAA	CACGCCAGCT	GCCAACAACC	GGGCCAGAAG	AGAAGCACCA	6000
CAGTACATGA	ACTACACAAT	CAATACCACA	AAAAACCTAA	ATGTATCAAT	AAGCAAGAAA	6060
AGGAAACGAA	GATTTCTGGG	CTTCTTGTTA	GGTGTAGGAT	CTGCAATAGC	AAGTGGTATA	6120
GCTGTATCCA	AAGTTTTACA	CCTTGAAGGA	GAAGTGAACA	AAATCAAAAA	TGCTTTGTTG	6180
TCTACAAACA	AAGCTGTAGT	CAGTCTATCA	AATGGGGTCA	GTGTTTTAAC	CAGCAAAGTG	6240
TTAGATCTCA	AGAATTACAT	AARTARCCGA	ATATTACCCA	TAGTAAATCA	ACAGAGCTGT	6300
CGCATCTCCA	ACATTGAAAC	AGTTATAGAA	TTCCAGCAGA	AGAATAGCAG	ATTGTTGGAA	6360
ATCACCAGAG	AATTTAGTGT	TAATGCAGGT	GTAACAACAC	CTTTAAGCAC	TTACATGTTA	6420
ACAAACAGTG	AGTTACTATC	ATTGATCAAT	GATATGCCTA	TAACAAATGA	CCAGAAAAA	6480
TTAATGTCAA	GCAATGTTCA	GATAGTAAGG	CAACAAAGTT	ATTCTATCAT	GTCTATAATA	6540
AAGGAAGAAG	TCCTTGCATA	TGTTGTACAG	CTACCTATCT	ATGGTGTAAT	AGATACACCT	6600
TGCTGGAAAT	TACACACATC	ACCTCTATGC	ACCACCAACA	TCAAAGAAGG	ATCAAATATT	6660
TGTTTAACA	GGACTGATAG	AGGATGGTAT	TGTGATAATG	CAGGATCAGT	ATCCTTCTTC	6720
CCACAGGCTG	ATACTTGCAA	AGTACAGTCC	AATCGAGTAT	TTTGTGACAC	TATGAACAGT	6780
TTAACATTAC	CAAGTGAAGT	CAGCCTTTGT	AACACTGACA	TATTCAATTC	CAAGTATGAC	6840
TGCAAAATT	TGACATCAAA	AACAGACATA	AGCAGCTCAG	TAATTACTTC	TCTTGGAGCT	6900
ATAGTGTCAT	GCTATGGAAA	AACTAAATGC	ACTGCATCCA	ATAAAAATCG	TGGGATTATA	6960
AAGACATTT	CTAATGGTTG	TGACTATGTG	TCAAACAAAG	GAGTAGATAC	TGTGTCAGTG	7020
GGCAACACT	TATACTATGT	AAACAAGCTG	GAAGGCAAAA	ACCTTTATGT	AAAAGGGGAA	7080

- 315 -

CCTATAATA	A ATTACTATE	A TCCTCTAGT	G TTTCCTTCT	G ATGAGTTTG	A TGCATCAATA	7140
TCTCAAGTC	A ATGAAAAA:	r Caatcaaag	T TTAGCTTTT	A TTCGTAGAT	C TGATGAATTA	7200
CTACATAAT	G TAAATACTG	G CAAATCTAC	r acaaatatt	A TGATAACTA	C AATTATTATA	7260
GTAATCATT	G TAGTATTGT	r atcattaati	A GCTATTGGT	TACTGTTGT	A TTGCAAAGCC	7320
AAAAACACA	C CAGTTACACT	T AAGCAAAGA	CAACTAAGT	GAATCAATA	A TATTGCATTC	7380
AGCAAATAG	A CAAAAAACTI	CTTAATCAT	TTTCAACAA	AATCTGCTG	A CCACCAATCC	7440
CAAATCAAC	TAACAACAA	A TATTTCAAC!	A TCATAGCACA	GGCTGAATC	TTTCCTCATA	7500
TCATGCTAC	CTACACAACTA	AGCTAGATCT	TCAACTCATA	GTTACATAA	AACCCCAAGT	7560
ATCACAATC	AACACTAAAT	CGACACATCA	TTCACAAAAT	TAACAACTGG	GGCAAATATG	7620
TCGCGAAGAI	ATCCTTGTAR	ATTTGAGATT	AGAGGTCATI	GCTTGAATGG	TAGAAGATGT	7680
CACTACAGTO	: ATAATTATTI	TGAATGGCCT	CCTCATGCAT	TACTAGTGAG	GCAAAACTTC	7740
ATGTTAAACA	AGATACTTAA	GTCAATGGAC	AAAAGCATAG	ACACTTTGTC	GGAAATAAGT	7800
GGAGCTGCTG	AACTGGATAG	AACAGAAGAA	TATGCTCTTG	GTATAGTTGG	AGTGCTAGAG	7860
AGTTACATAG	GATCAATAAA	CAACATAACA	AAACAATCAG	CATGTGTTGC	TATGAGTAAA	7920
CTTCTTATTG	AGATCAACAG	TGATGACATT	AAAAAACTGA	GAGATAACGA	AGAACCCAAT	7980
TCGCCTAAGA	TAAGAGTGTA	CAATACTGTT	ATATCATACA	TTGAGAGCAA	TAGAAAAAC	8040
AACAAGCAAA	CCATCCATCT	GCTCAAAAGA	CTACCAGCAG	ACGTGCTGAA	GAAGACAATA	8100
AAGAACACAT	TAGATATCCA	CAAAAGCATA	ACCATAAGCA	ACTCAAAAGA	GTCAACCGTG	8160
AATGATCAAA	ATGACCAAAC	CAAAAATAAT	GATATTACCG	GATAAATATC	CTTGTAGTAT	8220
ATCATCCATA	TTGATTTCAA	GTGAAAGCAT	GATTGCTACA	TTCAATCATA	AAAACATATT	8280
ACAATTTAAC	CATAACCATT	TGGATAACCA	CCAGTGTTTA	TTAAATCATA	TATTTGATGA	8340
AATTCATTGG	ACACCTAAAA	ACTTATTAGA	TGCCACTCAA	CAATTTCTCC	AACATCTTAA	8400
CATCCCTGAA	GATATATATA	CAGTATATAT	ATTAGTGTCA	TAATGCTTGA	CCATAACAAT	8460
TTTATATCAT	TCAACCATAA	AACAACCTTA	ATAAGGTTAT	GGGACAAAAT	GGATCCCATT	8520
					TGTTATCTCT	8580
TTTTCAGAAT	GTAATGCTTT	AGGGAGTTAC	CTTTTTAACG	GCCCCTATCT	TAAAAATGAT	8640

- 316 -

TACACCAACT	TAATTAGTAG	ACAAAGCCCA	CTACTAGAGC	ATATGAATCT	AAAAAAACTA	8700
ACTATAACAC	AGTCATTAAT	ATCTAGATAT	CATAAAGGTG	AACTGAAGTT	AGAAGAACCA	8760
ACTTATTTCC	AGTCATTACT	TATGACATAT	AAAAGTATGT	CCTCGTCTGA	ACAAATTGCT	8820
ACAACTAATT	TACTTAAAAA	AATAATACGA	AGAGCTATAG	AAATAAGTGA	TGTAAAGGTG	8880
TACGCCATCT	TGAATAAACT	GGGACTAAAG	GAAAAGGACA	GAGTTAAGCC	CAACAATAAT	8940
TCAGGTGATG	AAAACTCAGT	TCTTACAACC	ATAATCAAAG	ATGATATACT	TTCAGCTGTG	9000
GAAAACAATC	AATCATATAC	AAATTCAGAC	AAAAATCATT	CAGTAAATCA	AAATATCACT	9060
ATCAAAACAA	CACTCTTGAA	AAAATTGATG	TGTTCAATGC	AACATCCTCC	ATCATGGTTA	9120
ATACACTGGT	TCAATTTATA	TACAAAATTA	AATAACATAT	TAACACAATA	TCGATCAAAT	9180
GAGGTAAAA	GTCATGGGTT	TATATTAATA	GATAATCAAA	CTTTAAGTGA	TTTTCAGTTT	9240
ATTTTAAATC	AATATGGTTG	TATCGTTTAT	CATAAAGGAC	TCAAAAAAAT	CACAACTACT	9300
ACTTACAATC	AATTTTTGAC	ATGGAAAGAC	ATCAGCCTTA	GCAGATTAAA	TGTTTGCTTA	9360
ATTACTTGGA	TAAGTAATTG	TTTAAATACA	TTAAATAAAA	GCTTAGGGCT	GAGATGTGGA	9420
TTCAATAATG	TTGTGTTATC	ACAACTATTT	CTTTATGGAG	ATTGTATACT	GAAATTATTC	9480
CATAATGAAG	GCTTCTACAT	AATAAAAGAA	GTAGAGGGAT	TTATTATGTC	TTTAATTCTA	9540
AACATAACAG	AAGAAGATCA	ATTTAGGAAA	CGATTTTATA	ATAGCATGCT	AAATAACATC	9600
ACAGATGCAG	CTATTAAGGC	TCAAAAAAAC	CTACTATCAA	GAGTATGTCA	CACTTTATTA	9660
GACAAGACAG	TGTCTGATAA	TATCATAAAT	GGTAAATGGA	TAATCCTATT	AAGTAAATTT	9720
CTTAAATTGA	TTAAGCTTGC	AGGTGATAAT	AATCTCAATA	ACTTGAGTGA	GCTTTATTTT	9780
CTCTTCAGAA	TCTTTGGACA	TCCAATGGTC	GATGAAAGAC	AAGCAATGGA	TGCTGTAAGA	9840
ATTAACTGTA	ATGAAACCAA	GTTCTACTTA	TTAAGTAATC	TAAGTACGTT	AAGAGGTGCT	9900
TTCATTTATA	GAATCATAAA	GGGGTTTGTA	AATACCTACA	ACAGATGGCC	CACTTTAAGG	9960
AATGCTATTG	TTCTACCTCT	AAGATGGTTG	AACTATTATA	AACTTAATAC	TTATCCATCT	10020
CTACTTGAAA	TCACAGAGAA	AGATTTGATT	ATTTTATCAG	GATTGCGGTT	CTATCGTGAG	10080
TTTCATCTGC	CTAAAAAAGI	GGATCTTGAA	ATGATAATAA	ATGACAAAGC	CATTTCACCT	10140
CCAAAAGATT	TAATATGGAC	TAGTTTTCCT	AGAAATTACA	TGCCATCACA	TATACAAAAT	10200

- 317 -

TATATAGAAC	ATGAAAAGTT	GAAGTTCTCT	GAAAGTGACA	GATCAAGAAG	AGTACTAGAG	10260
TATTACTTGA	GAGATAATAA	ATTCAATGAA	TGCGATCTAT	' ACAATTGTGT	GGTCAATCAA	10320
AGCTATCTCA	ACAACTCTAA	CCATGTGGTA	TCACTAACTG	GTAAAGAAAG	AGAGCTCAGT	10380
GTAGGTAGAA	TGTTTGCTAT	GCAACCAGGT	ATGTTTAGGC	AAATTCAAAT	CTTAGCAGAG	10440
AAAATGATAG	CCGAAAATAT	TTTACAATTC	TTCCCTGAGA	GTTTGACAAG	ATATGGTGAT	10500
CTAGAGCTTC	AAAAGATATT	AGAATTAAAA	GCAGGAATAA	GCAACAAGTC	AAATCGTTAT	10560
AATGATAACT	ACAACAATTA	TATCAGTAAA	TGTTCTATCA	TTACAGACCT	TAGCAAATTC	10620
AATCAAGCAT	TTAGATATGA	AACATCATGT	ATCTGCAGTG	ATGTATTAGA	TGAACTGCAT	10680
GGAGTACAAT	CTCTGTTCTC	TTGGTTGCAT	TTAACAATAC	CTCTTGTCAC	AATAATATGT	10740
ACATATAGAC	ATGCACCTCC	TTTTATAAAG	GATCATGTTG	TTAATCTTAA	TAAAGTTGAT	10800
GAACAAAGTG	GATTATACAG	ATATCATATG	GGTGGTATTG	AAGGCTGGTG	TCAAAAACTG	10860
TGGACCATTG	AAGCTATATC	ATTATTAGAT	CTAATATCTC	TCAAAGGGAA	ATTCTCTATC	10920
ACAGCTCTAA	TAAATGGTGA	TAATCAGTCA	ATTGATATAA	GTAAACCAGT	TAGACTTATA	10980
GAGGGTCAGA	CCCATGCTCA	AGCAGATTAT	TTGTTAGCAT	TAAATAGCCT	TAAATTGCTA	11040
		AGGCCACAAG				11100
		AACAATCCAG				11160
•		TCCATGGATA				11220
		AACACAGGAG				11280
		TTGGTTATAC				11340
		ATATTTAGAT				11400
		TGATATGGCT				11460
					TCCAGACTTC	11520
					TCACGATTTA	11580
					ATGTATCATC	11640
					ACAGGCTTTA	11700
GGGTCTGAAA	GGCAAGCAAA	AATTACTAGT	GAGATTAATA	GATTAGCAGT	GACAGAAGTC	11760

- 318 -

TTAAGTATAG	CTCCAAACAA	AATATTTTCT	AAAAGTGCAC	AACATTATAC	TACCACTGAG	11820
ATTGATCTAA	ATGATATTAT	GCAAAATATA	GAACCAACTT	ACCCTCATGG	ATTAAGAGTT	11880
GTTTATGAAA	GTTTACCTTT	TTATAAAGCA	GAAAAAATAG	TTAATCTTAT	ATCAGGAACA	11940
AAATCCATAA	CTAATATACT	TGAAAAAACA	TCAGCAATAG	ATTCAACTGA	TATTAATAGG	12000
GCTACTGATA	TGATGAGGAA	AAATATAACT	TTACTTATAA	GGATACTTCC	ACTAGATTGT	12060
AACAAAGACA	AAAGAGAGTT	ATTAAGTTTA	GAAAATCTTA	GTATAACTGA	ATTAAGCAAG	12120
TATGTAAGAG	AAAGATCTTG	GTCGTTATCC	AATATAGTAG	GAGTAACATC	GCCAAGTATT	12180
ATGTTCACAA	TGGACATTAA	ATATACAACT	AGCACTATAG	CCAGTGGTAT	AATTATAGAA	12240
AAATATAATG	TTAATAGTTT	AACTCGTGGT	GAAAGAGGAC	CTACTAAGCC	ATGGGTAGGT	12300
TCATCTACGC	AGGAGAAAA	AACAATGCCA	GTGTACAATA	GACAAGTTTT	AACCAAAAAG	12360
CAAAGAGACC	AAATAGATTT	ATTAGCAAAA	TTAGACTGGG	TATATGCATC	CATAGACAAC	12420
AAAGATGAAT	TCATGGAAGA	ACTGAGTACT	GGAACACTTG	GACTGTCATA	TGAGAAAGCC	12480
AAAAATTGT	TTCCACAATA	TCTAAGTGTC	AATTATTTAC	ACCGCTTAAC	AGTCAGTAGT	12540
AGACCATGTG	AATTCCCTGC	ATCAATACCA	GCTTATAGAA	CAACAAATTA	TCATTTCGAT	12600
ACTAGTCCTA	TCAACCATGT	ATTAACAGAA	aagtatggag	ATGAAGATAT	CGACATTGTG	12660
TTTCAAAATT	GCATAAGTTT	TGGTCTTAGC	TTAATGTCGG	TTGTGGAACA	ATTCACAAAC	12720
ATATGTCCTA	ATAGAATTAT	TCTCATACCG	AAGCTGAATG	AGATACATTT	GATGAAACCT	12780
CCTATATTTA	CAGGAGATGT	TGATATCATC	AAGTTGAAGC	AAGTGATACA	AAAACAGCAC	12840
ATGTTCCTAC	CAGATAAAAT	AAGTTTAACC	CAATATGTAG	AATTATTCCT	AAGTAACAAA	12900
GCACTTAAAT	CTGGATCTCA	CATCAACTCT	AATTTAATAT	TAGTACATAA	AATGTCTGAT	12960
TATTTTCATA	ATGCTTATAT	TTTAAGTACT	AATTTAGCTG	GACATTGGAT	TCTGATTATT	13020
CAACTTATGA	AGGATTCAAA	AGGTATTTT	GAAAAAGATT	GGGGAGAGGG	GTATATAACT	13080
GATCATATG	TCATTAATTT	GAATGTTTTC	TTTAATGCTT	ATAAGACTTA	TTTGCTATGT	13140
TTTCATAAA	GTTATGGTAA	AGCAAAATTA	GAATGTGATA	TGAACACTTC	AGATCTTCTT	13200
TGTGTTTTG	AGCTAATAGA	CAGTAGCTAC	TGGAAATCTA	TGTCTAAAGT	TTTCCTAGAA	13260
CAAAAAGTC	TAAAATACAT	AATCAATCAA	GACACAAGTI	TGCATAGAAT	AAAAGGTTGT	13320

- 319 -

CATAGTTTTA AGTTATGGTT TTTAAAACGC CTTAATAATG CTAAATTTAC CGTATGCCCT	13380
TGGGTTGTTA ACATAGATTA TCACCCAACA CACATGAAAG CTATATTATC TTACATAGAT	13440
TTAGTTAGAA TGGGGTTAAT AAATGTAGAT AAATTAACCA TTAAAAATAA AAATAAATTC	13500
AATGATGAAT TTTACACATC AAATCTCTTT TACATTAGTT ATAACTTTTC AGATAACACT	13560
CATTTGCTAA CAAAACAAAT AAGAATTGCT AATTCAGAAT TAGAAAATAA TTATAACAAA	13620
CTATATCACC CAACCCCAGA AACTTTAGAA AATATGTCAT TAATTCCTGT CAAAAGTAAT	13680
AATAGTAATA AACCTAAATT TGGTATAAGT GGAAATACCG AATCTATGAT GACGTCAACA	13740
TTCTCCAATA AAACGCATAT TAAATCTTCC GCTGTTATTA CAAGATTCAA TTATAGTAAA	13800
CAAGACTTGT ACAATTTATT TCCAATTGTC GTGATAGACA GGATTATAGA TCATTCAGGT	13860
AATACAGCAA AATCTAACCA ACTCTACACT ACCACTTCAC ATCAGACATC TTTAGTAAGG	13920
AATAGTGCAT CACTTTATTG CATGCTTCCT TGGCATCATG TCAATAGATT TAACTTTGTA	13980
TTTAGTTCCA CAGGATGCAA GATCAGTATA GAGTATATTT TAAAAGATCT TAAGATTAAA	14040
GACCCCAGTT GTATAGCATT CATAGGTGAA GGAGCTGGTA ACTTATTATT ACGTACAGTA	14100
GTAGAACTTC ATCCAGACAT AAGATACATT TACAGAAGTT TAAAAGATTG CAATGATCAT	14160
AGTTTACCTA TTGAATTTCT AAGGTTATAC AACGGGCATA TAAACATAGA TTATGGTGAG	14220
AATTTAACCA TTCCTGCTAC AGATGCAACT AATAACATTC ATTGGTCTTA TTTACATATA	14280
ARATTTGCAG AACCTATTAG CATTTTTGTC TGCGATGCTG AATTACCTGT TACAGCCAAT	14340
TGGAGTAAAA TTATAATTGA ATGGAGTAAG CATGTAAGAA AGTGCAAGTA CTGTTCCTCT	14400
GTAAATAGAT GCATTTTAAT TGCAAAATAT CATGCCCAAG ATGATATTGA TTTCAAATTA	14460
GATAACATTA CTATATTAAA AACTTACGTG TGCCTAGGTA. GCAAGTTAAA AGGATCTGAA	14520
GTTTACTTAG TCCTTACAAT AGGCCCTGCA AATATACTTC CTGTTTTTAA TGTTGTGCAA	14580
AATGCTAAAT TGATTCTTTC AAGGACTAAA AATTTCATTA TGCCTAAAAA AACTGACAAA	14640
GAATCTATCG ATGCAAATAT TAAAAGCTTA ATACCTTTCC TTTGTTACCC TATAACAAAA	14700
AAAGGAATTA AGACTTCATT GTCAAAATTG AAGAGTGTAG TTAGTGGAGA TATATTATCA	14760
TATTCTATAG CTGGACGTAA TGAAGTATTC AGCAACAAGC TTATAAACCA CAAGCATATG	14820
AATATCCTAA AATGGCTAGA TCATGTTTTA AACTTTAGAT CAGCTGAACT TAATTACAAT	14880

- 320 -

CATTTATATA	TGATAGAGTC	CACATATCCT	TACTTAAGTG	AATTGTTAAA	CAGTTTAACA	14940
ACCAATGAGC	TCAAGAAGCT	GATTAAAATA	ACAGGTAGTG	TACTATACAA	CCTTCCCAAC	15000
GAACAGTAAC	TTAAAACATC	ATTARCAAGT	TTGATCAAAT	TTAGATGCTA	ACACATCATA	15060
ATATTATAGT	TATTAAAAA	TATATATGCA	AACTTTTCAA	TAATTTAGCA	TATTGATTCC	15120
AAAGTTATCA	TTTTGGTCTT	AAGGGGTTGA	ATAAAAATCT	AAAACTAACA	ATTATACATG	15180
TGCATTTACA	ACACAACGAG	ACATTAGTTT	TTGACACTTT	TTTTCTCGT		15229

### (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15

Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly 20 25 30

Ser Tyr Leu Phe Asn Gly Pro Tyr Leu Lys Asn Asp Tyr Thr Asn Leu 35 40 45

Ile Ser Arg Gln Ser Pro Leu Leu Glu His Met Asn Leu Lys Lys Leu 50 55 60

Thr Ile Thr Gln Ser Leu Ile Ser Arg Tyr His Lys Gly Glu Leu Lys 65 70 75 80

Leu Glu Glu Pro Thr Tyr Phe Gln Ser Leu Leu Met Thr Tyr Lys Ser 85 90 95

Met Ser Ser Glu Gln Ile Ala Thr Thr Asn Leu Leu Lys Lys Ile 100 105 110

Ile Arg Arg Ala Ile Glu Ile Ser Asp Val Lys Val Tyr Ala Ile Leu 115 120 125

- 321 -

	Ası	130	s Le O	u Gl	y Lei	u Lys	135		a Ası	) Ar	y Vai	140		Ası	a Ası	n As
	Se:	G1 <sub>3</sub>	ү Ав	p Gl	u Asr	150		. Lev	Thr	Th:	11e		Lys	a Asp	AB]	9 Il 16
	Leu	ı Se	r Al	a Va	165		. Asn	Gln	Sez	170		: Asn	Sez	. Yel	17!	
				1 Ası 180	0				185					190	)	
			19:					200					205	i		
		210	)	r Tha			215					220				
	225			s Ser		230					235					240
				n Phe	245					250					255	
				260	•				265					270		
			275					280					285			
		290		Leu			295					300				
	305			Val		310					315					320
				Phe	325					330					335	
				Met 340					345					350		
			355					360					365			
		370		Gln			375					380				
	385			Val		390					395					400
•	-cu	SAL	TAR	Phe	ren	гÀв	ren	TTO	rys	Leu	Ala	Gly	qeA	Asn	Asn	Leu

- 322 -

				405					410					415	
Asn	Asn	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
Met	Val	Asp 435	Glu	Arg	Gln	Ala	Met 440	Asp	Ala	Val	Arg	Ile 445	Asn	Сув	Asn
Glu	Thr 450	Lys	Phe	Tyr	Leu	Leu 455	Ser	Asn	Leu	Ser	Thr 460	Leu	Arg	Gly	Ala
Phe 465	Ile	Tyr	Arg	Ile	Ile 470	Lys	Gly	Phe	Val	Asn 475	Thr	Tyr	Asn	Arg	Trp 480
Pro	Thr	Leu	Arg	Asn 485	Ala	Ile	Val	Leu	Pro 490	Leu	Arg	Trp	Leu	Asn 495	Tyr
Tyr	Lys	Leu	Asn 500	Thr	Tyr	Pro	Ser	Leu 505	Leu	Glu	Ile	Thr	Glu 510	Lys	Asp
		515					520					525	His		
_	530					535					540		Ile		
545					550					555			Met		560
				565					570				Ser	575	
			580					585					Asn 590		
		595					600					605			
	610					615	;				620		Glu		
625					630					635			Gln		640
				645	•				650	)			Phe	655	,
			660	)				665	•				670	•	
Leu	Ly:	Ala 679		, Ile	Sez	. Ası	1 Lys		Asr	Arg	туг	685	a Asp	Asn	Туг

Asn	690		r Ile	Ser	Lys	695		: Ile	ı Ile	Th	700		ı Ser	Ly	3 Ph
Asn 705		. Ala	A Phe	Arg	710		Thr	Ser	Сув	715		S Sez	As <sub>r</sub>	Val	1 Le
Asp	Glu	. Leu	His	725		Gln	Ser	Leu	730		Trp	Leu	Hie	735	
			740					745	•				750	I	
		755					760					765	İ		
	770		Tyr			775					780				
785			Glu		790					795					800
			Ile	805					810					815	
			Pro 820					825					830		
		835					840					845			
	850		Gly			855					860				
865			Phe		870					875					880
			Asp	885					890					895	
			900 Glu					905					910		
		915	Trp				920					925			
	930		His			935					940				
945		-, <u></u>			950		+1-	200	vaħ	955	neu	nys	val	nen	960

- His Leu Lys Thr Phe Phe Asn Leu Asp Ser Ile Asp Met Ala Leu Thr 965 970 975
- Leu Tyr Met Asn Leu Pro Met Leu Phe Gly Gly Asp Pro Asn Leu
  980 985 990
- Leu Tyr Arg Ser Phe Tyr Arg Arg Thr Pro Asp Phe Leu Thr Glu Ala 995 1000 1005
- Ile Val His Ser Val Phe Val Leu Ser Tyr Tyr Thr Gly His Asp Leu 1010 1015 1020
- Gln Asp Lys Leu Gln Asp Leu Pro Asp Asp Arg Leu Asn Lys Phe Leu 1025 1030 1035 1040
- Thr Cys Ile Ile Thr Phe Asp Lys Asn Pro Asn Ala Glu Phe Val Thr 1045 1050 1055
- Leu Met Arg Asp Pro Gln Ala Leu Gly Ser Glu Arg Gln Ala Lys Ile 1060 1065 1070
- Thr Ser Glu Ile Asn Arg Leu Ala Val Thr Glu Val Leu Ser Ile Ala 1075 1080 1085
- Pro Asn Lys Ile Phe Ser Lys Ser Ala Gln His Tyr Thr Thr Glu 1090 1095 1100
- Ile Asp Leu Asn Asp Ile Met Gln Asn Ile Glu Pro Thr Tyr Pro His 1105 1110 1115 1120
- Gly Leu Arg Val Val Tyr Glu Ser Leu Pro Phe Tyr Lys Ala Glu Lys 1125 1130 1135
- Ile Val Asn Leu Ile Ser Gly Thr Lys Ser Ile Thr Asn Ile Leu Glu 1140 1145 1150
- Lys Thr Ser Ala Ile Asp Ser Thr Asp Ile Asn Arg Ala Thr Asp Met 1155 1160 1165
- Met Arg Lys Asn Ile Thr Leu Leu Ile Arg Ile Leu Pro Leu Asp Cys 1170 1175 1180
- Asn Lys Asp Lys Arg Glu Leu Leu Ser Leu Glu Asn Leu Ser Ile Thr 1185 1190 1195 1200
- Glu Leu Ser Lys Tyr Val Arg Glu Arg Ser Trp Ser Leu Ser Asn Ile 1205 1210 1215
- Val Gly Val Thr Ser Pro Ser Ile Met Phe Thr Met Asp Ile Lys Tyr 1220 1225 1230
- Thr Thr Ser Thr Ile Ala Ser Gly Ile Ile Glu Lys Tyr Asn Val

- 325 -

1235		1240	1245	
Asn Ser Leu Th 1250	or Arg Gly Glo 12	u Arg Gly Pr	o Thr Lys Pro Tr 1260	o Val Gly
Ser Ser Thr Gl 1265	n Glu Lys Ly: 1270	s Thr Met Pro	o Val Tyr Asn Arg 1275	g Gln Val 1280
Leu Thr Lys Ly	s Gln Arg Asp 1285	Gln Ile Asp 129	o Leu Leu Ala Lys 90	Leu Asp 1295
Trp Val Tyr Al 13	a Ser Ile Asp 00	Asn Lys Ası 1305	o Glu Phe Met Glu 131	
Ser Thr Gly Th 1315	r Leu Gly Leu	ı Ser Tyr Glu 1320	Lys Ala Lys Lys 1325	Leu Phe
1330	133	5	Arg Leu Thr Val	
1342	1350		Ala Tyr Arg Thr 1355	1360
	1365	137		1375
138	30	1385	Asn Cys Ile Ser 1396	0
1395		1400	Thr Asn Ile Cys 1405	
1410	1415	5	Ile His Leu Met 1420	
1425	1430		Lys Leu Lys Gln 1435	1440
	1445	1450		1455
146	0	1465	Lys Ser Gly Ser 1470	
14/5		1480	Ser Asp Tyr Phe : 1485	
1490	1495		His Trp Ile Leu : 1500	
Gln Leu Met Lys 1505	Asp Ser Lys (		Glu Lys Asp Trp ( 1515	Gly Glu 1520

- 326 -

- Gly Tyr Ile Thr Asp His Met Phe Ile Asn Leu Asn Val Phe Phe Asn 1525 1530 1535
- Ala Tyr Lys Thr Tyr Leu Leu Cys Phe His Lys Gly Tyr Gly Lys Ala 1540 1545 1550
- Lys Leu Glu Cys Asp Met Asn Thr Ser Asp Leu Leu Cys Val Leu Glu 1555 1560 1565
- Leu Ile Asp Ser Ser Tyr Trp Lys Ser Met Ser Lys Val Phe Leu Glu 1570 1575 1580
- Gln Lys Val Ile Lys Tyr Ile Ile Asn Gln Asp Thr Ser Leu His Arg 1585 1590 1595 1600
- Ile Lys Gly Cys His Ser Phe Lys Leu Trp Phe Leu Lys Arg Leu Asn 1605 1610 1615
- Asn Ala Lys Phe Thr Val Cys Pro Trp Val Val Asn Ile Asp Tyr His 1620 1625 1630
- Pro Thr His Met Lys Ala Ile Leu Ser Tyr Ile Asp Leu Val Arg Met 1635 1640 1645
- Gly Leu Ile Asn Val Asp Lys Leu Thr Ile Lys Asn Lys Asn Lys Phe 1650 1655 1660
- Asn Asp Glu Phe Tyr Thr Ser Asn Leu Phe Tyr Ile Ser Tyr Asn Phe 1665 1670 1675 1686
- Ser Asp Asn Thr His Leu Leu Thr Lys Gln Ile Arg Ile Ala Asn Ser 1685 1690 1695
- Glu Leu Glu Asn Asn Tyr Asn Lys Leu Tyr His Pro Thr Pro Glu Thr 1700 1705 1710
- Leu Glu Asn Met Ser Leu Ile Pro Val Lys Ser Asn Asn Ser Asn Lys 1715 1720 1725
- Pro Lys Phe Gly Ile Ser Gly Asn Thr Glu Ser Met Met Thr Ser Thr 1730 1740
- Phe Ser Asn Lys Thr His Ile Lys Ser Ser Ala Val Ile Thr Arg Phe 1745 1750 1755 1760
- Asn Tyr Ser Lys Gln Asp Leu Tyr Asn Leu Phe Pro Ile Val Val Ile 1765 1770 1775
- Asp Arg Ile Ile Asp His Ser Gly Asn Thr Ala Lys Ser Asn Gln Leu 1780 1785 1790

- 327 -

- Tyr Thr Thr Ser His Gln Thr Ser Leu Val Arg Asn Ser Ala Ser 1795 1800 1805
- Leu Tyr Cys Met Leu Pro Trp His His Val Asn Arg Phe Asn Phe Val 1810 1815 1820
- Phe Ser Ser Thr Gly Cys Lys Ile Ser Ile Glu Tyr Ile Leu Lys Asp 1825 1830 1835 1840
- Leu Lys Ile Lys Asp Pro Ser Cys Ile Ala Phe Ile Gly Glu Gly Ala 1845 1850 1855
- Gly Asn Leu Leu Arg Thr Val Val Glu Leu His Pro Asp Ile Arg 1860 1865 1870
- Tyr Ile Tyr Arg Ser Leu Lys Asp Cys Asn Asp His Ser Leu Pro Ile 1875 1880 1885
- Glu Phe Leu Arg Leu Tyr Asn Gly His Ile Asn Ile Asp Tyr Gly Glu 1890 1895 1900
- Asn Leu Thr Ile Pro Ala Thr Asp Ala Thr Asn Asn Ile His Trp Ser 1905 1910 1915 1920
- Tyr Leu His Ile Lys Phe Ala Glu Pro Ile Ser Ile Phe Val Cys Asp 1925 1930 1935
- Ala Glu Leu Pro Val Thr Ala Asn Trp Ser Lys Ile Ile Ile Glu Trp 1940 1945 1950
- Ser Lys His Val Arg Lys Cys Lys Tyr Cys Ser Ser Val Asn Arg Cys 1955 1960 1965
- Ile Leu Ile Ala Lys Tyr His Ala Gln Asp Asp Ile Asp Phe Lys Leu 1970 1975 1980
- Asp Asn Ile Thr Ile Leu Lys Thr Tyr Val Cys Leu Gly Ser Lys Leu 1985 1990 1995 2000
- Lys Gly Ser Glu Val Tyr Leu Val Leu Thr Ile Gly Pro Ala Asn Ile 2005 2010 2015
- Leu Pro Val Phe Asn Val Val Gln Asn Ala Lys Leu Ile Leu Ser Arg 2020 2025 2030
- Thr Lys Asn Phe Ile Met Pro Lys Lys Thr Asp Lys Glu Ser Ile Asp 2035 2040 2045
- Ala Asn Ile Lys Ser Leu Ile Pro Phe Leu Cys Tyr Pro Ile Thr Lys 2050 2055 2060
- Lys Gly Ile Lys Thr Ser Leu Ser Lys Leu Lys Ser Val Val Ser Gly

PCT/US97/16718 WO 98/13501

- 328 -

	2065			2070				2075				2080					
	Asp	Ile	Leu	Ser	Tyr 208	Ser	Ile	Ala	Gly	Arg 2090		Glu	Val	Phe	Ser 209		
	Lys	Leu	Ile	Asn 2100		Lys	His	Met	Asn 210		Leu	Lys	Trp	Leu 211		His	
	Val	Leu	Asn 211		Arg	Ser	Ala	Glu 212		Asn	Tyr	Asn	His 212		Tyr	Met	
	Ile	Glu 2130		Thr	Туг	Pro	Tyr 213		Ser	Glu	Leu	Leu 214		Ser	Leu	Thr	
	Thr 214		Glu	Leu	Lys	Lys 215		Ile	Lys	Ile	Thr 215		Ser	Val	Leu	Tyr 2160	
	Asn	Leu	Pro	Asn	Glu 216												
(2)	INFO	RMAT:	ION 1	FOR :	SEQ	ID N	0:27	:									
	(i)	(A (B (C	) LE ) TY ) ST	ngth Pe: : Rand:	: 15 nucl EDNE	TERI: 219   eic SS: line	base acid sing	pai	rs								
	(ii)	MOL	ECUL	E TY	PB:	RNA	(gen	omic	)								
	(xi)	SEO	UENC	E DE	SCRI	PTIO	N: S	EQ I	D NO	:27:							
ACGG	GAAA	A AA	ATGC	GTAC	T AC	AAAC	TTGC	ACA	TTCG	AAA	AAAA	TGGG	GC A	AATA	AGAA	С	6
TTGA	TAAG	TG C	TATT	TAAG	т ст	AACC	TTTT	CAA	TCAG	AAA	TGGG	GTGC	AA T	TCAC	TGAG	C	12
ATGA	AAAT	GG T	TAGA	TTAC	AA A	ATTT	ATTT	GAC	AATG	ACG	aagt	agca	TT G	AATT	TAAA	A	18
ACAT	GTTA	TA C	TGAT	TAAA	T AA	TTCT	TCTG	ACC	DTAA	CAT	TAGC	CAAA	GC A	GCAA	TACA	T	24

ACAATTAAAT TAAACGGCAT AGTTTTTATA CATGTTATAA CAAGCAGTGA AGTGTGCCCT

GATAACAATA TTGTAGTGAA ATCTAACTTT ACAACAATGC CAATACTACA AAATGGAGGA

TACATATGGG AATTGATTGA GTTGACACAC TGCTCTCAAT TAAACGGTTT AATGGATGAT

AATTGTGAAA TCAAATTTTC TAAAAGACTA AGTGACTCAG TAATGACTAA TTATATGAAT

CARATATCTG ACTTACTTGG GCTTGATCTC AATTCATGAA TTATGTTTAG TCTAATTCAA

300

360

420

480

540

- 329 -

TAGACATGTG TTTATTACCA TTTTAGTTAA TATAAAAACT CATCAAAGGG AAATGGGGCA	600
ANTANACTCA CCTANTCANT CANACCATGA GCACTACANA TGACANCACT ACTATGCANA	660
GATTGATGAT CACAGACATG AGACCCCTGT CAATGGATTC AATAATAACA TCTCTTACCA	720
AAGAAATCAT CACACAAA TTCATATACT TGATAAACAA TGAATGTATT GTAAGAAAAC	780
TTGATGAAAG ACAAGCTACA TTTACATTCT TAGTCAATTA TGAGATGAAG CTACTGCACA	840
AAGTAGGGAG TACCAAATAC AAAAAATACA CTGAATATAA TACAAAATAT GGCACTTTCC	900
CCATGCCTAT ATTTATCAAT CACGGCGGGT TTCTAGAATG TATTGGCATT AAGCCTACAA	960
AACACACTCC TATAATATAC AAATATGACC TCAACCCGTG AATTCCAACA AAAAAACCAA	1020
CCCAACCAAA CCAAACTATT CCTCAAACAA CAGTGCTCAA TAGTTAAGAA GGAGCTAATC	1080
CATTTTAGTA ATTAAAAATA AAAGTAAAGC CAATAACATA AATTGGGGCA AATACAAAGA	1140
TGGCTCTTAG CAAAGTCAAG TTGAATGATA CATTAAATAA GGATCAGCTG CTGTCATCCA	1200
GCAAATACAC TATTCAACGT AGTACAGGAG ATAATATTGA CACTCCCAAT TATGATGTGC	1260
AAAAACACCT AAACAAACTA TGTGGTATGC TATTAATCAC TGAAGATGCA AATCATAAAT	1320
TCACAGGATT AATAGGTATG TTATATGCTA TGTCCAGGTT AGGAAGGGAA GACACTATAA	1380
AGATACTTAA AGATGCTGGA TATCATGTTA AAGCTAATGG AGTAGATATA ACAACATATC	1440
GTCAAGATAT AAATGGAAAG GAAATGAAAT TCGAAGTATT AACATTATCA AGCTTGACAT	1500
CAGAAATACA AGTCAATATT GAGATAGAAT CTAGAAAGTC CTACAAAAAA ATGCTAAAAG	1560
AGATGGGAGA AGTGGCTCCA GAATATAGGC ATGATTCTCC AGACTGTGGG ATGATAATAC	1620
TGTGTATAGC TGCACTTGTG ATAACCAAAT TAGCAGCAGG AGACAGATCA GGTCTTACAG	1680
CAGTAATTAG GAGGGCAAAC AATGTCTTAA AAAACGAAAT AAAACGATAC AAGGGCCTCA	1740
TACCAAAGGA TATAGCTAAC AGTTTTTATG AAGTGTTTGA AAAACACCCT CATCTTATAG	1800
ATGTTTTCGT GCACTTTGGC ATTGCACAAT CATCCACAAG AGGGGGTAGT AGAGTTGAAG	1860
GAATCTTTGC AGGATTGTTT ATGAATGCCT ATGGTTCAGG GCAAGTAATG CTAAGATGGG	1920
GAGTTTTAGC CAAATCTGTA AAAAATATCA TGCTAGGACA TGCTAGTGTC CAGGCAGAAA	1980
TGGAGCAAGT TGTGGAAGTC TATGAGTATG CACAGAAGTT GGGAGGAGAA GCTGGATTCT	2040
ACCATATATT GAACAATCCA AAAGCATCAT TGCTGTCATT AACTCAATTT CCCAACTTCT	2100

- 330 -

CAAGTGTGGT	CCTAGGCAAT	GCAGCAGGTC	TAGGCATAAT	GGGAGAGTAT	AGAGGTACAC	2160
CAAGAAACCA	GGATCTTTAT	GATGCAGCTA	AAGCATATGC	AGAGCAACTC	AAAGAAAATG	2220
GAGTAATAAA	CTACAGTGTA	TTAGACTTAA	CAGCAGAAGA	ATTGGAAGCC	ATAAAGCATC	2280
AACTCAACCC	CAAAGAAGAT	GATGTAGAGC	TTTAAGTTAA	CARARATAC	GGGGCAAATA	2340
AGTCAACATG	GAGAAGTTTG	CACCTGAATT	TCATGGAGAA	GATGCAAATA	ACAAAGCTAC	2400
CAAATTCCTA	GAATCAATAA	AGGGCAAGTT	CGCATCATCC	AAAGATCCTA	AGAAGAAAGA	2460
TAGCATAATA	TCTGTTAACT	CAATAGATAT	AGAAGTAACT	AAAGAGAGCC	CGATAACATC	2520
TGGCACCAAC	ATCATCAATC	CAACAAGTGA	AGCCGACAGT	ACCCCAGAAA	CAAAAGCCAA	2580
CTACCCAAGA	AAACCCCTAG	TAAGCTTCAA	AGAAGATCTC	ACCCCAAGTG	ACAACCCTTT	2640
TTCTAAGTTG	TACAAGGAAA	CAATAGAAAC	ATTTGATAAC	aatgaagaag	AATCTAGCTA	2700
CTCATATGAA	GAGATAAATG	ATCAAACAAA	TGACAACATT	ACAGCAAGAC	TAGATAGAAT	2760
TGATGAAAA	TTAAGTGAAA	TATTAGGAAT	GCTCCATACA	TTAGTAGTTG	CAAGTGCAGG	2820
ACCCACTTCA	GCTCGCGATG	GAATAAGAGA	TGCTATGGTT	GGTCTAAGAG	AAGAGATGAT	2880
AGAAAAAATA	AGAGCGGAAG	CATTAATGAC	CAATGATAGG	TTAGAGGCTA	TGGCAAGACT	2940
TAGGAATGAG	GAAAGCGAAA	AAATGGCAAA	AGACACCTCA	GATGAAGTGT	CTCTTAATCC	3000
AACTTCCAAA	AAATTGAGTG	ACTTGTTGGA	AGACAACGAT	AGTGACAATG	ATCTATCACT	3060
TGATGATTTT	TGATCAGCGA	TCAACTCACT	CAGCAATCAA	CAACATCAAT	AAAACAGACA	3120
TCAATCCATT	GAATCAACTG	CCAGACCGAA	CAAACAAACG	TCCATCAGTA	GAACCACCAA	3180
CCAATCAATC	AACCAATTGA	TCAATCAGCA	ACCCGACAAA	ATTAACAATA	TAGTAACAAA	3240
AAAAGAACAA	GATGGGGCAA	ATATGGAAAC	ATACGTGAAC	AAGCTTCACG	AAGGCTCCAC	3300
ATACACAGCA	GCTGTTCAGT	ACAATGTTCT	AGAAAAAGAT	GATGATCCTG	CATCACTAAC	3360
AATATGGGTG	CCTATGTTCC	AGTCATCTGT	GCCAGCAGAC	TTGCTCATAA	AAGAACTTGC	3420
AAGCATCAAT	ATACTAGTGA	AGCAGATCTC	TACGCCCAAA	GGACCTTCAC	TACGAGTCAC	3480
					TAAGCGCAAA	
					GTGAAATCAA	
AGCATGCAGT	CTAACATGCT	TAAAAGTAAA	AAGTATGTTA	ACTACAGTCA	AAGATCTTAC	3660

- 331 -

CATGAAGAC	A TTCAACCCC	A CTCATGAGA	T CATTGCTCT	A TGTGAATTT	AAAATATTAT	3720
GACATCAAA	A AGAGTAATA	A TACCAACCT	A TCTAAGATC	A ATTAGTGTC	A AGAACAAGGA	3780
TCTGAACTC	A CTAGAAAAT	A TAGCAACCA	C CGAATTCAA	A AATGCTATC	CCAATGCAAA	3840
AATTATTCC	T TATGCAGGA	I TAGTGTTAG	TATCACAGT:	r actgacaata	AAGGAGCATT	3900
CAAATATAT	C AAACCACAGI	A GTCAATTTA:	r agtagatet	GGTGCCTACC	TAGAAAAAGA	3960
GAGCATATA	T TATGTGACT	CTAATTGGAL	A GCATACAGC	CACGTTTT	CAATCAAACC	4020
ACTAGAGGA	T TAAACTTAA1	TATCAACAC	r Gaatgacago	F TCCACATATA	TCCTCAAACT	4080
ACACACTAT.	A TCCAAACATC	ATAAACATCI	T ACACTACACA	CTTCATCACA	CAAACCAATC	4140
CCACTCAAA	A TCCAAAATCA	CTACCAGCC	CTATCCGCTA	GACCTAGAGT	GCGAATAGGC	4200
AAATAAAAC	C AAAATATGGG	GTAAATAGAC	ATTAGTTAGA	GTTCAATCAA	TCTTAACAAC	4260
CATTTATAC	C GCCAATTCAA	CACATATACI	ATAAATCTTA	AAATGGGAAA	TACATCCATC	4320
ACAATAGAA	C TCACAAGCAA	ATTTTGGCCC	TATTTTACAC	TARTACATAT	GATCTTAACT	4380
CTAATCTTT	TACTAATTAT	AATCACTATO	ATGATTGCAA	CACTAAATAA	GCTAAGTGAA	4440
CACAAAGCAT	TCTGCAACAA	AACTCTTGAA	CTAGGACAGA	TGTACCAAAT	CAACACACAG	4500
AGTTCCACCA	TTATGCTGTG	TCAAACCATA	ATCCTGTATA	TACAAACAAA	CAAATCCAAT	4560
CCTCTCACAG	AGTCACGGTG	TCGCAAAACC	ACGCTAACCA	TCATGGTAGC	ATAGAGTAGT	4620
<b>EAAAAATTTA</b>	TTAACATAAT	GATGAATTGT	TAGTATGAGA	TCAAAAACAA	CATTGGGGCA	4680
AATGCAACCA	TGTCCAAACA	CAAGAATCAA	CGCACTGCCA	GGACTCTAGA	AAAGACCTGG	4740
GATACTCTTA	ATCATCTAAT	TGTAATATCC	TCTTGTTTAT	ACAGATTAAA	TTTAAAATCT	4800
ATAGCACAAA	TAGCACTATC	AGTTTTGGCA	ATGATAATCT	CAACCTCTCT	CATAATTGCA	4860
GCCATAATAT	TCATCATCTC	TGCCAATCAC	AAAGTTACAC	TAACAACGGT	CACAGTTCAA	4920
ACAATAAAA	ACCACACTGA	AAAAAACATC	ACCACCTACC	CTACTCAAGT	CTCACCAGAA	4980
AGGGTTAGTT	CATCCAAGCA	ACCCACAACC	ACATCACCAA	TCCACACAAG	TTCAGCTACA	5040
ACATCACCCA	ATACAAAATC	AGAAACACAC	CATACAACAG	CACAAACCAA	AGGCAGAACC	5100
ACCACTTCAA	CACAGACCAA	CAAGCCAAGC	ACAAAACCAC	GTCCAAAAAA	TCCACCAAAA	5160
<b>Aaagatgatt</b>	ACCATTTTGA	AGTGTTCAAC	TTCGTTCCCT	GCAGTATATG	TGGCAACAAT	5220

- 332 -

CAACTTTGCA AATCCATCTG CAAAACAATA CCAAGCAACA AACCAAAGAA GAAACCAACC	5280
ATCARACCCA CARACARACC RACCACCARA ACCACARACA ARAGAGACCC ARARACACCA	5340
GCCARACGA CGRARAAGA AACTACCACC AACCCAACAA AAAAACTAAC CCTCAAGACC	5400
ACAGAAAGAG ACACCAGCAC CTCACAATCC ACTGCACTCG ACACAACCAC ATTAAAACAC	5460
ACAGTCCAAC AGCAATCCCT CCTCTCAACC ACCCCCGAAA ACACACCCAA CTCCACACAA	5520
ACACCCACAG CATCCGAGCC CTCCACACCA AACTCCACCC AAAAAACCCA GCCACATGCT	5580
TAGTTATICA AAAACTACAT CTTAGCAGAG AACCGTGATC TATCAAGCAA GAACGAAATT	5640
AAACCTGGGG CAAATAACCA TGGAGTTGAT GATCCACAAG TCAAGTGCAA TCTTCCTAAC	5700
TCTTGCTATT AATGCATTGT ACCTCACCTC AAGTCAGAAC ATAACTGAGG AGTTTTACCA	5760
ATCGACATGT AGTGCAGTTA GCAGAGGTTA TTTTAGTGCT TTAAGAACAG GTTGGTATAC	5820
TAGTGTCATA ACAATAGAAT TAAGTAATAT AAAAGAAACC AAATGCAATG GAACTGACAC	5880
TAAAGTAAAA CTTATGAAAC AAGAATTAGA TAAGTATAAG AATGCAGTAA CAGAATTACA	5940
GCTACTTATG CAAAACACAC CAGCTGTCAA CAACCGGGCC AGAAGAGAAG	6000
TATGAACTAC ACAATCAATA CCACTAAAAA CCTAAATGTA TCAATAAGCA AGAAGAGGAA	6060
ACGAAGATTT CTAGGCTTCT TGTTAGGTGT GGGATCTGCA ATAGCAAGTG GTATAGCTGT	6120
ATCARARGIT CTACACCITG ARGGAGARGI GARCARGAIC ARARAIGCII IGIIGICTAC	6180
AAACAAAGCT GTAGTCAGTT TATCAAATGG GGTCAGTGTT TTAACCAGCA AAGTGTTAGA	6240
TCTCAAGAAT TACATAAATA ACCAATTATT ACCCATAGTA AATCAACAGA GCTGTCGCAT	6300
CTCCAACATT GAAACAGTTA TAGAATTCCA GCAGAAGAAC AGCAGATTGT TGGAAATCAC	6360
CAGAGAATTT AGTGTCAATG CAGGTGTAAC AACACCTTTA AGCACTTACA TGTTGACAAA	6420
CAGTGAGTTA CTATCATTAA TCAATGATAT GCCTATAACA AATGATCAGA AAAAATTAAT	6480
GTCAAGCAAT GTTCAGATAG TAAGGCAACA AAGTTATTCC ATCATGTCTA TAATAAAGGA	6540
AGAAGTCCTT GCATATGTTG TACAGCTGCC TATCTATGGT GTAATAGATA CACCTTGCTG	6600
GAAATTGCAC ACATCGCCTC TATGCACTAC CAACATCAAA GAAGGATCAA ATATTTGTTT	
AACAAGGACT GATAGAGGAT GGTATTGTGA TAATGCAGGA TCAGTATCCT TCTTTCCACA	
GGCTGACACT TGTARAGTAC AGTCCARTCG AGTATTTTGT GACACTATGA ACAGTTTGAC	6780

- 333 -

ATTACCAAGT	GAAGTCAGCC	TTTGTAACAC	TGACATATTC	AATTCCAAGT	ATGACTGCAA	6840
AATTATGACA	TCAAAAACAG	ACATAAGCAG	CTCAGTAATT	ACTTCTCTTG	GAGCTATAGT	6900
GTCATGCTAT	GGTAAAACTA	AATGCACTGC	ATCCAACAAA	AATCGTGGGA	TTATAAAGAC	6960
ATTTTCTAAT	GGTTGTGACT	ATGTGTCAAA	CAAAGGAGTA	GATACTGTGT	CAGTGGGCAA	7020
CACTITATAC	TATGTAAACA	AGCTGGAAGG	CAAGAACCTT	TATGTAAAAG	GGGAACCTAT	7080
AATAAATTAC	TATGACCCTC	TAGTGTTTCC	TTCTGATGAG	TTTGATGCAT	CAATATCTCA	7140
AGTCAATGAA	AAAATCAATC	AAAGTTTAGC	TTTTATTCGT	AGATCTGATG	AATTACTACA	7200
TAATGTAAAT	ACTGGCAAAT	CTACTACAAA	TATTATGATA	ACTACAATTA	TTATAGTAAT	7260
CATTGTAGTA	TTGTTATCAT	TAATAGCTAT	TGGTTTACTG	TTGTATTGTA	AAGCCAAAAA	7320
CACACCAGTT	ACACTAAGCA	AAGACCAACT	AAGTGGAATC	AATAATATTG	CATTCAGCAA	7380
ATAGACAAAA	AACCACCTGA	TCATGTTTCA	ACAACAATCT	GCTGACCACC	AATCCCAAAT	7440
CAACTTACAA	CAAATATTTC	AACATCACAG	TACAGGCTGA	ATCATTTCCT	CACATCATGC	7500
TACCCACATA	ACTAAGCTAG	ATCCTTAACT	TATAGTTACA	TAAAAACCTC	AAGTATCACA	7560
ATCAACCACT	AAATCAACAC	ATCATTCACA	AAATTAACAG	CTGGGGCAAA	TATGTCGCGA	7620
AGAAATCCTT	GTAAATTTGA	GATTAGAGGT	CATTGCTTGA	ATGGTAGAAG	ATGTCACTAC	7680
AGTCATAATT	ACTITGAATG	GCCTCCTCAT	GCATTACTAG	TGAGGCAAAA	CTTCATGTTA	7740
AACAAGATAC	TCAAGTCAAT	GGACAAAAGC	ATAGACACTT	TGTCTGAAAT	AAGTGGAGCT	7800
GCTGAACTGG	ATAGAACAGA	AGAATATGCT	CTTGGTATAG	TTGGAGTGCT	AGAGAGTTAC	7860
ATAGGATCTA	TAAACAACAT	AACAAAACAA	TCAGCATGTG	TTGCTATGAG	TAAACTTCTT	7920
ATTGAGATCA	ATAGTGATGA	CATTAAAAAG	CTTAGAGATA	ATGAAGAACC	CAATTCACCT	7980
AAGATAAGAG	TGTACAATAC	TGTTATATCA	TACATTGAGA	GCAATAGAAA	AAACAACAAG	8040
CARACCATCC	ATCTGCTCAA	GAGACTACCA	GCAGACGTGC	TGAAGAAGAC	AATAAAGAAC	8100
ACATTAGATA	TCCACAAAAG	CATAACCATA	AGCAATCCAA	AAGAGTCAAC	TGTGAATGAT	8160
CAAAATGACC	AAACCAAAAA	TAATGATATT	ACCGGATAAA	TATCCTTGTA	GTATATCATC	8220
CATATTGATC	TCAAGTGAAA	GCATGGTTGC	TACATTCAAT	CATAAAAACA	TATTACAATT	8280
TAACCATAAC	TATTTGGATA	ACCACCAGCG	TTTATTAAAT	CATATATTTG	ATGAAATTCA	8340

- 334 -

TTGGACACCT	AAAAACTTAT	TAGATGCCAC	TCAACAATTT	CTCCAACATC	TTAACATCCC	8400
TGAAGATATA	TATACAGTAT	ATATATTAGT	GTCATAATGC	TTGACCATAA	CGACTCTATG	8460
TCATCCAACC	ATAAAACTAT	TTTGATAAGG	TTATGGGACA	AAATGGATCC	CATTATTAAT	8520
GGAAACTCTG	CTAATGTGTA	TCTAACTGAT	AGTTATTTAA	AAGGTGTTAT	CTCTTTTCA	8580
GAGTGTAATG	CTTTAGGGAG	TTATCTTTTT	AACGGCCCTT	ATCTTAAAAA	TGATTACACC	8640
AACTTAATTA	GTAGACAAAG	CCCACTACTA	GAGCATATGA	ATCTTAAAAA	ACTAACTATA	8700
ACACAGTCAT	TAATATCTAG	ATATCATAAA	GGTGAACTGA	AATTAGAAGA	ACCAACTTAT	8760
TTCCAGTCAT	TACTTATGAC	ATATAAAAGT	ATGTCCTCGT	CTGAACAAAT	TGCTACAACT	8820
AACTTACTTA	TAAAAAAT	ACGAAGAGCC	ATAGAAATAA	GTGATGTAAA	GGTGTACGCC	8880
ATCTTGAATA	AACTAGGATT	AAAGGAAAAG	GACAGAGTTA	AGCCCAACAA	TAATTCAGGT	8940
GATGAAAACT	CAGTACTTAC	AACCATAATT	AAAGATGATA	TACTTTCGGC	TGTGGAAAAC	9000
AATCAATCAT	ATACAAATTC	AGACAAAAGT	CACTCAGTAA	ATCAAAATAT	CACTATCAAA	9060
ACAACACTCT	TGAAAAAATT	GATGTGTTCA	ATGCAACATC	CTCCATCATG	GTTAATACAC	9120
TGGTTCAATT	TATATACAAA	ATTAAATAAC	ATATTAACAC	AATATCGATC	AAATGAGGTA	9180
AAAAGTCATG	GGTTTATATT	AATAGATAAT	CAAACTTTAA	GTGGTTTTCA	GTTTATTTTA	9240
AATCAATATG	GTTGTATCGT	TTATCATAAA	GGACTCAAAA	AAATCACAAC	TACTACTTAC	9300
AATCAATTTT	TGACATGGAA	AGACATCAGC	CTTAGCAGAT	TAAATGTTTG	CTTAATTACT	9360
TGGATAAGTA	ATTGTTTAAA	TACATTAAAC	AAAAGCTTAG	GGCTGAGATG	TGGATTCAAT	9420
AATGTTGTGT	TATCACAATT	ATTTCTTTAT	GGAGATTGTA	TACTGAAATT	ATTTCATAAT	9480
GAAGGCTTCT	ACATAATAAA	AGAAGTAGAG	GGATTTATTA	TGTCTTTAAT	TCTAAACATA	9540
ACAGAAGAAG	ATCAATTTAA	GAAACGATTT	TATAATAGCA	TGCTAAATAA	CATCACAGAT	9600
GCAGCTATTA	AGGCTCAAAA	GGACCTACTA	TCAAGAGTAT	GTCACACTTT	ATTAGACAAG	9660
ACAGTGTCTG	ATAATATCAT	AAATGGTAAA	TGGATAATCC	TATTAAGTAA	ATTTCTTAAA	9720
TTGATTAAGC	TTGCAGGTGA	TAATAATCTC	AATAACTTGA	GTGAGCTATA	TTTTCTCTTC	9780
AGAATCTTTG	GACATCCAAT	GGTCGATGAA	AGACAAGCAA	TGGATTCTGT	AAGAATTAAC	9840
TGTAATGAAA	CTAGGTTCTA	CTTATTAAGT	AGTCTAAGTA	CATTAAGAGG	TGCTTTCATT	9900

- 335 -

TATAGAATCA	TAAAAGGGTT	TGTAAATACO	TACAACAGAT	GGCCCACCTI	AAGGAATGCT	9960
ATTGTCCTAC	CTCTAAGATG	GTTAAACTAC	TATAAACTTA	ATACTTATCO	ATCTCTACTT	10020
GAAATCACAG	AAAATGATTT	GATTATTTA	TCAGGATTGC	GGTTCTATCG	TGAGTTTCAT	10080
CTGCCTAAAA	AAGTGGATCT	TGAAATGATA	ATAAATGACA	AAGCCATTTC	ACCTCCAAAA	10140
GATCTAATAT	GGACTAGTTT	TCCTAGAAAT	TACATGCCAT	CACATATACA	AAATTATATA	10200
GAACATGAAA	AGTTGAAGTT	CTCTGAAAGC	GACAGATCGA	GAAGAGTACT	AGAGTATTAC	10260
TTGAGAGATA	ATAAATTCAA	TGAATGCGAT	CTATACAATT	GTGTAGTCAA	TCAAAGCTAT	10320
CTCAACAACT	CTAATCACGT	GGTATCACTA	ACTGGTAAAG	AAAGAGAGCT	CAGTGTAGGT	10380
AGAATGTTTG	CTATGCAACC	AGGTATGTTT	AGGCAAATCC	AAATCTTAGC	AGAGAAAATG	10440
ATAGCTGAAA	ATATTTTACA	ATTCTTCCCT	GAGAGTTTGA	CAAGATATGG	TGATCTAGAG	10500
CTTCAAAAGA	TATTAGAATT	AAAAGCAGGA	ATAAGCAACA	AGTCAAATCG	TTATAATGAT	10560
AACTACAACA	ATTATATCAG	TAAATGTTCT	ATCATTACAG	ATCTTAGCAA	ATTCAATCAG	10620
GCATTTAGAT	ATGAAACATC	ATGTATCTGC	AGTGATGTAT	TAGATGAACT	GCATGGAGTA	10680
CAATCTCTGT	TCTCTTGGTT	GCATTTAACA	ATACCTCTTG	TCACAATAAT	ATGTACATAT	10740
AGACATGCAC	CTCCTTTCAT	AAAGGATCAT	GTTGTTAATC	TTAATGAGGT	TGATGAACAA	10800
AGTGGATTAT	ACAGATATCA	TATGGGTGGT	ATTGAGGGCT	GGTGTCAAAA	ACTGTGGACC	10860
ATTGAAGCTA	TATCATTATT	AGATCTAATA	TCTCTCAAAG	GGAAATTCTC	TATCACAGCT	10920
CTGATAAATG	GTGATAATCA	GTCAATTGAT	ATAAGCAAAC	CAGTTAGACT	TATAGAGGGT	10980
CAGACCCATG	CACAAGCAGA	TTATTTGTTA	GCATTAAATA	GCCTTAAATT	GTTATATAAA	11040
GAGTATGCAG	GTATAGGCCA	TAAGCTTAAG	GGAACAGAGA	CCTATATATC	CCGAGATATG	11100
CAGTTCATGA	GCAAAACAAT	CCAGCACAAT	GGAGTGTACT	ATCCAGCCAG	TATCAAAAA	11160
GTCCTGAGAG	TAGGTCCATG	GATAAACACG	ATACTTGATG	ATTTTAAAGT	TAGTTTAGAA	11220
TCTATAGGCA	GCTTAACACA	GGAGTTAGAA	TACAGAGGAG	AAAGCTTATT	ATGCAGTTTA	11280
ATATTTAGGA	ACATTTGGTT	ATACAATCAA	ATTGCTTTGC	AACTCCGAAA	TCATGCATTA	11340
TGTAACAATA	AGCTATATTT	AGATATATTG	AAAGTATTAA	AACACTTAAA	AACTTTTTT	11400
AATCTTGATA	GCATTGATAT	GGCTTTATCA	TTGTATATGA	ATTTGCCTAT	GCTGTTTGGT	11460

- 336 -

GGTGGTGATC	CTAATTTGTT	ATATCGAAGC	TTTTATAGGA	GAACTCCAGA	CTTCCTTACA	11520
GAAGCTATAG	TACATTCAGT	GTTTGTGTTG	AGCTATTATA	CTGGTCACGA	TTTACAAGAT	11580
AAGCTCCAGG	ATCTTCCAGA	TGATAGACTG	AACAAATTCT	TGACATGTGT	CATCACATTI'	11640
GATAAAAATC	CCAATGCCGA	GTTTGTAACA	TTGATGAGGG	ATCCACAGGC	TTTAGGGTCT	11700
GAAAGGCAAG	CTAAAATTAC	TAGTGAGATT	AATAGATTAG	CAGTAACAGA	AGTCTTAAGT	11760
ATAGCCCCAA	ACAAAATATT	TTCTAAAAGT	GCACAACATT	ATACTACCAC	TGAGATTGAT	11820
CTAAATGACA	TTATGCAAAA	TATAGAACCA	ACTTACCCTC	ATGGATTAAG	AGTTGTTTAT	11880
GAAAGTTTAC	CTTTTTATAA	AGCAGAAAAA	ATAGTTAATC	TTATATCAGG	AACAAAATCC	11940
ATAACTAATA	TACTTGAAAA	AACATCAGCA	ATAGATACAA	CTGATATTAA	TAGGGCTACT	12000
gatatgatga	GGAAAAATAT	AACTTTACTT	ATAAGGATAC	TTCCACTAGA	TTGTAACAAA	12060
GACAAAAGAG	AGTTATTAAG	TTTAGAAAAT	CTTAGTATAA	CTGAATTAAG	CAAGTATGTA	12120
agagaagat	CTTGGTCATT	ATCCAATATA	GTAGGAGTAA	CATCGCCAAG	TATTATGTTC	12180
ACAATGAACA	TTAAATATAC	AACTAGCACT	ATAGCCAGTG	GTATAATAAT	AGAAAATAT	12240
AATGTTAATA	GTTTAACTCG	TGGTGAAAGA	GGACCCACCA	AGCCATGGGT	AGGCTCATCC	12300
ACGCAGGAGA	AAAAAACAAT	GCCAGTGTAC	AACAGACAAG	TTTTAACCAA	AAAGCAAAGA.	12360
GACCAAATAG	ATTTATTAGC	AAAATTAGAC	TGGGTATATG	CATCCATAGA	CAACAAAGAT	12420
GAATTCATGG	AAGAACTGAG	TACTGGAACA	CTTGGACTGT	CATATGAAAA	AGCCAAAAAG	12480
TTGTTTCCAC	AATATCTAAG	TGTCAATTAT	TTACACCGTT	TAACAGTCAG	TAGTAGACCA	12540
TGTGAATTCC	CTGCATCAAT	ACCAGCTTAT	AGAACAACAA	ATTATCATTT	TGATACTAGT	12600
CCTATCAATC	ATGTATTAAC	AGAAAAGTAT	GGAGATGAAG	ATATCGACAT	TGTGTTTCAA.	12660
	GTTTTGGTCT					12720
CCTAATAGAA	TTATTCTCAT	ACCGAAGCTG	AATGAGATAC	ATTTGATGAA	ACCTCCTATA	12780
TTTACAGGAG	ATGTTGATAT	CATCAAGTTG	AAGCAAGTGA	TACAAAAGCA	GCACATGTTC	12840
CTACCAGATA	AAATAAGTTT	AACCCAATAT	GTAGAATTAT	TCTTAAGTAA	CAAAGCACTT'	12900
					TGATTATTT	
CATAATGCTT	ATATTTTAAG	TACTAATTTA	GCTGGACATT	GGATTCTGAT	TATTCAACTI	13020

- 337 -

ATGAAAGATT CAAAAGGTAT TTTTGAAAAA GATTGGGGAG AGGGGTACAT AACTGATCAT	13080
ATGTTCATTA ATTTGAATGT TTTCTTTAAT GCTTATAAGA CTTATTTGCT ATGTTTTCAT	13140
AAAGGTTATG GTAAAGCAAA ATTAGAATGT GATATGAACA CTTCAGATCT TCTTTGTGTT	13200
TTGGAGTTAA TAGACAGTAG CTACTGGAAA TCTATGTCTA AAGTTTTCCT AGAACAAAAA	13260
GTCATAAAAT ACATAGTCAA TCAAGACACA AGTTTGCGTA GAATAAAAGG CTGTCACAGT	13320
TTTAAGTTGT GGTTTTTAAA ACGCCTTAAT AATGCTAAAT TTACCGTATG CCCTTGGGTT	13380
GTTAACATAG ATTATCACCC AACACACATG AAAGCTATAT TATCTTACAT AGATTTAGTT	13440
AGAATGGGGT TAATAAATGT AGATAAATTA ACCATTAAAA ATAAAAACAA ATTCAATGAT	13500
GAATTTTACA CATCAAATCT CTTTTACATT AGTTATAACT TTTCAGACAA CACTCATTTG	13560
CTAACAAAAC AAATAAGAAT TGCTAATTCA GAATTAGAAG ATAATTATAA CAAACTATAT	13620
CACCCAACCC CAGAAACTTT AGAAAATATG TCATTAATTC CTGTTAAAAG TAATAATAGT	13680
AACAAACCTA AATTTTGTAT AAGTGGAAAT ACCGAATCTA TGATGATGTC AACATTCTCT	13740
AGTAAAATGC ATATTAAATC TTCCACTGTT ACCACAAGAT TCAATTATAG CAAACAAGAC	13800
TTGTACAATT TATTTCCAAT TGTTGTGATA GACAAGATTA TAGATCATTC AGGTAATACA	13860
GCAAAATCTA ACCAACTTTA CACCACCACT TCACATCAGA CATCTTTAGT AAGGAATAGT	13920
GCATCACTIT ATTGCATGCT TCCTTGGCAT CATGTCAATA GATTTAACTT TGTATTTAGT	13980
TCCACAGGAT GCAAGATCAG TATAGAGTAT ATTTTAAAAG ATCTTAAGAT TAAGGACCCC	14040
AGTTGTATAG CATTCATAGG TGAAGGAGCT GGTAACTTAT TATTACGTAC GGTAGTAGAA	14100
CTTCATCCAG ACATAAGATA CATTTACAGA AGTTTAAAAG ATTGCAATGA TCATAGTTTA	14160
CCTATTGAAT TTCTAAGGTT ATACAACGGG CATATAAACA TAGATTATGG TGAGAATTTA	14220
ACCATTCCTG CTACAGATGC AACTAATAAC ATTCATTGGT CTTATTTACA TATAAAATTT	14280
	14340
	14400
	14460
	14520
TTAATCCTTA CAATAGGCCC TGCAAATATA CTTCCTGTTT TTGATGTTGT ACAAAATGCT	14580

- 338 -

AAJ	<b>TTGATAC</b>	TTTCAAGAAC	TAAAAATTTC	ATTATGCCTA	AAAAAACTGA	CAAGGAATCT	14640
ATO	GATGCAA	ATATTAAAAG	CTTAATACCT	TTCCTTTGTT	ACCCTATAAC	AAAAAAAGGA	14700
AT]	TAAGACTT	CATTGTCAAA	ATTGAAGAGT	GTAGTTAATG	GAGATATATT	ATCATATTCT	14760
AT	AGCTGGAC	GTAATGAAGT	ATTCAGCAAC	AAGCTTATAA	ACCACAAGCA	TATGAATATC	14820
CT	<b>LAAATGG</b> C	TAGATCATGT	TTTAAATTTT	AGATCAGCTG	AACTTAATTA	CAATCATTTA	14880
TAC	CATGATAG	AGTCCACATA	TCCTTACTTA	AGTGAATTGT	TAAATAGTTT	AACAACCAA'I'	14940
GAG	CTCAAGA	AGCTGATTAA	AATAACAGGT	AGTGTGCTAT	ACAACCTTCC	CAACGAACAG	15000
TAC	STTTAAAA	TATCATTAAC	AAGTTTGGTC	AAATTTAGAT	GCTAACACAT	CATTATATTA	15060
TAC	TTATTAA	AGAATATACA	AACTTTTCAA	TAATTTAGCA	TATTGATTCC	AAAATTATCA	15120
TTI	TAGTCTT	AAGGGGTTAA	ATAAAAGTCT	AAAACTAACA	ATTATACATG	TGCATTCACA	15180
<b>AC</b>	CAACGAG	ACATTAGTTT	TTGACACTTT	TTTTCTCGT			15219

## (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15

Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly 20 25 30

Ser Tyr Leu Phe Asn Gly Pro Tyr Leu Lys Asn Asp Tyr Thr Asn Leu 35 40

Ile Ser Arg Gln Ser Pro Leu Leu Glu His Met Asn Leu Lys Lys Leu 50 55 60

Thr Ile Thr Gln Ser Leu Ile Ser Arg Tyr His Lys Gly Glu Leu Lys 65 70 75 80

- 339 -

Let	ı Glu	ı Gl	u Pro	Thr 85	Туг	Phe	e Glr	ı Sei	Let 90	u Le	n We	t Th	т Ту	r Ly: 95	s Ser
Met	: Sei	: Se	r Sei 100		Gln	Ile	a Ala	105		r Ası	a Let	ı Lei	11)		3 Ile
Ile	Arg	11!		lle	Glu	Ile	9 Ser 120		Va]	Lys	s Val	12!		a Ile	Leu
Asn	130	Let	ı Gly	Leu	Lys	Glu 135		Asp	Arg	y Val	140		AB1	n Asr	Asn
Ser 145	Gly	Ası	Glu	Asn	Ser 150	Val	. Leu	Thr	Thr	11e		Lys	a Asp	) Asp	11e 160
Leu	Ser	Ala	a Val	Glu 165	Asn	Asn	Gln	Ser	Tyr 170		Asn	Ser	Asp	Lys 175	
His	Ser	Val	180		Asn	Ile	Thr	Ile 185	Lys	Thr	Thr	Leu	190	_	Lys
		195					200					205			
	210		Thr			215					220				
225			Ser		230					235					240
			Phe	245					250					255	_
			Lys 260					265					270		_
		275	Ser				280					285		_	
	290		Leu			295		•			300				-
305			Val		310					315					320
			Phe	325					330					335	
Gly	Phe	Ile	Met 340	Ser :	Leu :	Ile		Asn 345	Ile	Thr	Glu	Glu	Asp 350	Gln	Phe

Lys	Lys	Arg 355	Phe	Tyr	Asn	Ser	Met 360	Leu	Asn	Asn	Ile	Thr 365	Asp	Ala	Ala
Ile	Lys 370	Ala	Gln	Lys	Asp	Leu 375	Leu	Ser	Arg	Val	Сув 380	His	Thr	Leu	Leu
Asp 385	Lys	Thr	Val	Ser	Asp 390	Asn	Ile	Ile	Asn	Gly 395	Lys	Trp	Ile	Ile	Leu 400
Leu	Ser	ГÀв	Phe	Leu 405	Lys	Leu	Ile	Lys	Leu 410	Ala	Gly	Asp	Asn	Авп 415	Leu
Asn	Asn	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
Met	Val	Авр 435	Glu	Arg	Gln	Ala	Met 440	Asp	Ser	Val	Arg	Ile 445	Asn	Сув	Asn
	Thr 450					455					460			•	
Phe 465	Ile	Tyr	Arg	Ile	11e 470	Lys	Gly	Phe	Val	Asn 475	Thr	Tyr	Asn	Arg	Trp 480
Pro	Thr	Leu	Arg	Asn 485	Ala	Ile	Val	Leu	Pro 490	Leu	Arg	Trp	Leu	Asn 495	Tyr
Tyr	Lys	Leu	Asn 500	Thr	Tyr	Pro	Ser	Leu 505	Leu	Glu	Ile	Thr	Glu 510	Asn	Авр
Leu	Ile	Ile 515	Leu	Ser	Gly	Leu	Arg 520	Phe	Tyr	Arg	Glu	Phe 525	His	Leu	?ro
Lys	<b>Lys</b> 530	Val	qaA	Leu	Glu	Met 535	Ile	Ile	Asn	qaA	Lys 540	Ala	Ile	Ser	Pro
Pro 545	Lys	Asp	Leu	Ile	Trp 550	Thr	Ser	Phe	Pro	Arg 555	Asn	Tyr	Met	Pro	3er 560
His	Ile	Gln	Asn	Tyr 565	Ile	Glu	His	Glu	Lys 570	Leu	Lys	Phe	Ser	Glu 575	3er
Asp	Arg	Ser	Arg 580	Arg	Val	Leu	Glu	Tyr 585	Tyr	Leu	Arg	<b>A</b> ap	Asn 590	Lys	?he
Asn	Glu	Сув 595	Asp	Leu	Tyr	Asn	Сув 600	Val	Val	Asn	Gln	Ser 605	Tyr	Leu	Asn
	Ser 610					615			•	•	620				
Val	Gly	Arg	Met	Phe	Ala	Met	Gln	Pro	Gly	Met	Phe	Arg	Gln	Ile	Gln

- 341 -

625					630					635					64
Ile	Leu	Ala	Glu	Lys 645	Met	Ile	Ala	Glu	Asn 650		Leu	Gln	Phe	Phe 655	
Glu	Ser	Leu	Thr 660		Tyr	Gly	Asp	Leu 665		Leu	Gln	Lys	11e		G1
Leu	Lys	Ala 675		Ile	Ser	Asn	Lys 680		Asn	Arg	Tyr	Asn 685	_	Авл	Ty
Asn	Asn 690	туг	Ile	Ser	Lys	Сув 695		Ile	Ile	Thr	<b>Asp</b> 700	Leu	Ser	Lys	Pho
Asn 705		Ala	Phe	Arg	Tyr 710	Glu	Thr	Ser	Cys	Ile 715	Сув	Ser	Asp	Val	Le:
Asp	Glu	Leu	His	Gly 725	Val	Gln	Ser	Leu	Phe 730	Ser	Trp	Leu	His	Leu 735	
Ile	Pro	Leu	Val 740	Thr	Ile	Ile	Сув	Thr 745	Tyr	Arg	His	Ala	Pro 750	Pro	Phe
		755			Val		760					765			
	770				Met	775					780				
785					Ile 790					795					800
•				805	Ala				810					815	
			820		Arg			825					830		
		835			Leu		840					845			
	850				Lys	855					860				
865					Ser 870					875					880
				885	Lys				890					895	
Ile	Leu	Авр	<b>Asp</b> 900	Phe	Lys	Val	Ser	Leu 905	Glu	Ser	Ile	Gly	Ser 910	Leu	Thr

- 342 -

Gln	Glu	Leu 915	Glu	Tyr	Arg	Gly	Glu 920	Ser	Leu	Leu	Сув	Ser 925	Leu	Ile	Phe
Arg	Asn 930	Ile	Trp	Leu	Tyr	Asn 935	Gln	Ile	Ala	Leu	Gln 940	Leu	Arg	Asn	His
Ala 945	Leu	Сув	Asn	Asn	<b>L</b> ув 950	Leu	Tyr	Leu	Asp	Ile 955	Leu	Lys	Val	Leu	<b>Lys</b> 960
His	Leu	Lys	Thr	Phe 965	Phe	Asn	Leu	Asp	Ser 970	Ile	Asp	Met	Ala	Leu 975	Ser
Leu	Tyr	Met	Asn 980	Leu	Pro	Met	Leu	Phe 985	Gly	Gly	Gly	yab	Pro 990	Asn	Leu
Leu	Tyr	Arg 995	Ser	Phe	Tyr	Arg	Arg 1000		Pro	qaA	Phe	Leu 1005		Glu	Ala
Ile	Val 101		Ser	Val	Phe	Val 1015		Ser	Tyr	Tyr	Thr 1020	_	His	Asp	Leu
Gln 102	_	Lys	Leu	Gln	Asp 1030	Leu )	Pro	Авр	Asp	Arg 1035		Asn	Lys	Phe	Leu 1040
Thr	Сув	Val	Ile	Thr 104		Asp	Lys	Asn	Pro 1050		Ala	Glu	Phe	Val 1055	
Leu	Met	Ara	Авр	Pro	Gln	Ala	Leu	Gly	Ser	Glu	Arg	Gln		Lys	Ile
		3	106	0				1065	;				1070	)	
Thr			106		Arg	Leu	Ala 1080	Val		Glu	Val	Leu 1085	Ser		Ala
	Ser	Glu 107! Lys	1060 Ile	Asn		Leu Lys 109	1080 Ser	Val	Thr			1085	Ser 5	Ile	
Pro	Ser Asn 109	Glu 107! Lys	1066 Ile 5	Asn	Ser	Lys 1099	1080 Ser	Val ) Ala	Thr	His	Tyr 1100	1085 Thr	Ser 5	Ile	Glu
Pro	Ser Asn 1096 Asp	Glu 107! Lys )	Ile Ile Ile Asn	Asn Phe Asp	Ser Ile 1110	Lys 1099	1080 Ser Gln	Val ) Ala Asn	Thr Gln Ile	His Glu 111!	Tyr 1100 Pro	1085 Thr )	Ser ; Thr Tyr	Ile Thr Pro	Glu His 1120 Lys
Pro	Asn 1096 Asp 5	Glu 107! Lys Leu Arg	Ile Ile Asn Val	Asn Phe Asp Val 112:	Ser Ile 1110 Tyr	Lys 1095 Met	Ser Ser Sln Ser	Val ) Ala Asn Leu	Thr Gln Ile Pro 1130 Ser	Glu 111! Phe	Tyr 1100 Pro ;	Thr Thr	Ser Thr Tyr	Thr Pro Glu 1135	Glu His 1120 Lys

Met Arg Lys Asn Ile Thr Leu Leu Ile Arg Ile Leu Pro Leu Asp Cys 1170 1175 1180

- 343 -

Asn 118		Asp	Lys	Arg	Glu 119		Leu	Ser	Leu	Glu 119		Leu	Ser	Ile	Thr 1200
Glu	Leu	Ser	Lys	Tyr 120		Arg	Glu	Arg	Ser 121		Ser	Leu	Ser	Asn 121	Ile 5
Val	Gly	Val	Thr 122		Pro	Ser	Ile	Met 122		Thr	Met	Asn	11e	_	Tyr
Thr	Thr	Ser 123		Ile	Ala	Ser	Gly 124		Ile	Ile	Glu	Lys 124		Asn	Val
Asn	Ser 125		Thr	Arg	Gly	Glu 125		Gly	Pro	Thr	Lys 126		Trp	Val	Gly
Ser 126		Thr	Gln	Glu	Lys 127		Thr	Met	Pro	Val 127		Asn	Arg	Gln	Val 1280
Leu	Thr	Lys	Lys	Gln 128		Asp	Gln	Ile	Asp 129		Leu	Ala	Lys	Leu 129	<b>дар</b> 5
Trp	Val	Tyr	Ala 130		Ile	Asp	Asn	Lys 130		Glu	Phe	Met	Glu 131		Leu
Ser	Thr	Gly 131		Leu	Gly	Leu	Ser 132	Tyr 0	Glu	Lys	Ala	Lys 132		Leu	Phe
Pro	Gln 133		Leu	Ser	Val	Asn 1335		Leu	His	Arg	Leu 134		Val	Ser	Ser
Arg 134		Сув	Glu	Phe	Pro 1350		Ser	Ile	Pro	Ala 135	_	Arg	Thr	Thr	Asn 1360
туг	His	Phe	Asp	Thr 1365		Pro	Ile	Asn	His 1370		Leu	Thr	Glu	Lys 1375	
Gly	Asp	Glu	Asp 1380		Asp	Ile	Val	Phe 1385		Asn	Сув	Ile	Ser 1390		Gly
Leu	Ser	Leu 1395		Ser	Val	Val	Glu 1400	Gln )	Phe	Thr	Asn	Ile 1405		Pro	Asn
	11e		Leu	Ile	Pro	Lys 1415		Asn	Glu	Ile	His 1420		Met	Lys	Pro
Pro 1425		Phe	Thr	Gly	Авр 1430		Asp	Ile	Ile	Lys 1435		Lys	Gln	Val	Ile 1440
Gln	Lys	Gln	His	Met	Phe	Leu	Pro	Asp	Lys	Ile	Ser	Leu	Thr	Gln	Tyr

Val Glu Leu Phe Leu Ser Asn Lys Ala Leu Lys Ser Gly Ser His Ile

1450

1445

- 344 -

			1460	)				1465	5				1470	)	
Asn	Ser	Asn 1475		Ile	Leu	Val	His 1480		Met	Ser	Asp	Tyr 1485		His	Asn
Ala	Tyr 1490		Leu	Ser	Thr	Asn 1495		Ala	Gly	His	Trp 1500		Leu	Ile	Ile
Gln 1505		Met	Lys	Asp	Ser 1510		Gly	Ile	Phe	Glu 1519	_	Asp	Trp	Gly	Glu 1520
Gly	Tyr	Ile	Thr	Авр 1525	His 5	Met	Phe	Ile	Asn 1530		Asn	Val	Phe	Phe 1535	
Ala	Tyr	Lys	Thr 1540		Leu	Leu	Сув	Phe 1545		Lys	Gly	Tyr	Gly 1550	_	Ala
Lys	Leu	Glu 1555	_	Asp	Met	Asn	Thr 1560		Asp	Leu	Leu	Сув 1565		Leu	Glu
Leu	Ile 1570	_	Ser	Ser	Tyr	Trp 1575	_	Ser	Met	Ser	Lys 1580		Phe	Leu	Glu
Gln 1585		Val	Ile	Lys	Tyr 1590		Val	Asn	Gln	Asp 159		Ser	Leu	Arg	Arg 1600
Ile	Lys	Gly	Сув	His 1605	Ser	Phe	Lys	Leu	Trp 1610		Leu	Lys	Arg	Leu 1615	
Asn	Ala	Lys	Phe 1620		Val	Сув	Pro	Trp 1625		Val	Asn	Ile	Asp 1630	_	His
Pro	Thr	His 1635		Lys	Ala	Ile	Leu 1640		Tyr	Ile	Asp	Leu 1645		Arg	Met
Gly	Leu 1650		Asn	Val	Авр	Lys 1655		Thr	Ile	Lys	Asn 1660	-	Asn	Lys	Phe
Asn 1665		Glu	Phe	Tyr	Thr 1670		Asn	Leu	Phe	Tyr 1675		Ser	Tyr	Asn	Phe 1680
Ser	Asp	Asn	Thr	Нів 1685	Leu	Leu	Thr	Lys	Gln 1690		Arg	Ile	Ala	Asn 1695	
Glu	Leu	Glu	Asp 1700		Tyr	Asn	Lys	Leu 170		His	Pro	Thr	Pro 1710		Thr
Leu	Glu	Asn 171		Ser	Leu	Ile	Pro 1720		Lys	Ser	Asn	Asn 1725		Asn	Lys
Pro	Lys 1730		Сув	Ile	Ser	Gly 1735		Thr	Glu	Ser	Met 1740		Met	Ser	Thr

- 345 -

Phe	Ser	Ser	Lys	Met	His	Ile	Lys	Ser	Ser	Thr	Val	Thr	Thr	Arg	Phe
1745					1750	)				175	5				1760

- Asn Tyr Ser Lys Gln Asp Leu Tyr Asn Leu Phe Pro Ile Val Val Ile 1765 1770 1775
- Asp Lys Ile Ile Asp His Ser Gly Asn Thr Ala Lys Ser Asn Gln Leu 1780 1785 1790
- Tyr Thr Thr Ser His Gln Thr Ser Leu Val Arg Asn Ser Ala Ser 1795 1800 . 1805
- Leu Tyr Cys Met Leu Pro Trp His His Val Asn Arg Phe Asn Phe Val 1810 1815 1820
- Phe Ser Ser Thr Gly Cys Lys Ile Ser Ile Glu Tyr Ile Leu Lys Asp 1825 1830 1835 1840
- Leu Lys Ile Lys Asp Pro Ser Cys Ile Ala Phe Ile Gly Glu Gly Ala 1845 1850 1855
- Gly Asn Leu Leu Arg Thr Val Val Glu Leu His Pro Asp Ile Arg 1860 1865 1870
- Tyr Ile Tyr Arg Ser Leu Lys Asp Cys Asn Asp His Ser Leu Pro Ile 1875 1880 1885
- Glu Phe Leu Arg Leu Tyr Asn Gly His Ile Asn Ile Asp Tyr Gly Glu 1890 1895 1900
- Asn Leu Thr Ile Pro Ala Thr Asp Ala Thr Asn Asn Ile His Trp Ser 1905 1910 1915 1920
- Tyr Leu His Ile Lys Phe Ala Glu Pro Ile Ser Ile Phe Val Cys Asp 1925 1930 1935
- Ala Glu Leu Pro Val Thr Ala Asn Trp Ser Lys Ile Ile Ile Glu Trp 1940 1945 1950
- Ser Lys His Val Arg Lys Cys Lys Tyr Cys Ser Ser Val Asn Arg Cys 1955 1960 1965
- Ile Leu Ile Ala Lys Tyr His Ala Gln Asp Asp Ile Asp Phe Lys Leu 1970 1975 1980
- Asp Asn Ile Thr Ile Leu Lys Thr Tyr Val Cys Leu Gly Ser Lys Leu 1985 1990 1995 2000
- Lys Gly Ser Glu Val Tyr Leu Ile Leu Thr Ile Gly Pro Ala Asn Ile 2005 2010 2015

- 346 -

	Leu	Pro	Val	Phe 2020	_	Val	Val	Gln	Asn 2025		Lys	Leu	Ile	Leu 2030		Arg	
	Thr	Lys	Asn 2035		Ile	Met	Pro	Lys 2040	_	Thr	Asp	Lys	Glu 2045	Ser	Ile	Asp	
	Ala	Asn 2050		Lys	Ser	Leu	Ile 2055		Phe	Leu	Сув	Tyr 2060		Ile	Thr	Lys	
	Lys 2065	-	Ile	Lys	Thr	Ser 207		Ser	Lys	Leu	Lys 2075		Val	Val	Asn	Gly 2080	
	Asp	Ile	Leu	Ser	Tyr 208!		Ile	Ala	Gly	Arg 2090		Glu	Val	Phe	Ser 2095		
	Lys	Leu	Ile	Asn 210		Lys	His	Met	Asn 210		Leu	Lys	Trp	Leu 2110		His	
	Val	Leu	Asn 211		Arg	Ser	Ala	Glu 212		Asn	Tyr	Asn	His 212	Leu 5	Tyr	Met	
	Ile	Glu 213		Thr	Tyr	Pro	Tyr 213		Ser	Glu	Leu	Leu 214		Ser	Leu	Thr	
	Thr 214		Glu	Leu	Lys	Lys 215		Ile	Lys	Ile	Thr 215	_	Ser	Val	Leu	Tyr 2160	
	Asn	Leu	Pro	Asn	Glu 216												
(2) 1	NFO	RMAT:	ION I	FOR S	BBQ :	ID N	0:29	:									
	(i)	(A (B (C	) LE ) TY: ) ST	e chi ngth pe: 1 rand! polo	: 15: nucle EDNE:	219   mic   SS:	base acid sing	pai	rs								
(	(ii)	MOL	ECUL	E TY	PE: 1	RNA	(gen	omic	)								
	( <b>x</b> i)	·SEO	UBNC	E DE:	SCRI:	PTIO	N: S	BQ II	D NO	:29:							
ACGG											AAAA'	rggg	GC A	LATA	AGAA	2	60
TTGAT	raag'	IG C	TATT	TAAG	r ct.	AACC'	TTTT	CAA	TCAG	AAA !	rggg	GTGC	AA T	rcac:	rgago	2	120
ATGAT	CAAA!	GG T	TAGA	TTAC	A AA	ATTT.	ATTT	GAC	DTAA	ACG 2	aagt	AGCA:	TT G	PTAA!	LTAAF	A	180

ACATGTTATA CTGATAAATT AATTCTTCTG ACCAATGCAT TAGCCAAAGC AGCAATACAT

240

- 347 -

ACAATTAAAT TAAACGGCAT AGTTTTTATA CATGTTATAA CAAGCAGTGA AGTGTGCCCT	300
GATAACAATA TTGTAGTGAA ATCTAACTTT ACAACAATGC CAATACTACA AAATGGAGGA	360
TACATATGGG AATTGATTGA GTTGACACAC TGCTCTCAAT TAAACGGTTT AATGGATGAT	420
AATTGTGAAA TCAAATTTTC TAAAAGACTA AGTGACTCAG TAATGACTAA TTATATGAAT	480
CAAATATCTG ACTTACTTGG GCTTGATCTC AATTCATGAA TTATGTTTAG TCTAATTCAA	540
TAGACATGTG TTTATTACCA TTTTAGTTAA TATAAAAACT CATCAAAGGG AAATGGGGCA	600
AATAAACTCA CCTAATCAAT CAAACCATGA GCACTACAAA TGACAACACT ACTATGCAAA	660
GATTGATGAT CACAGACATG AGACCCCTGT CAATGGATTC AATAATAACA TCTCTTACCA	720
AAGAAATCAT CACACAAAA TTCATATACT TGATAAACAA TGAATGTATT GTAAGAAAAC	780
TTGATGAAAG ACAAGCTACA TTTACATTCT TAGTCAATTA TGAGATGAAG CTACTGCACA	840
AAGTAGGGAG TACCAAATAC AAAAAATACA CTGAATATAA TACAAAATAT GGCACTTTCC	900
CCATGCCTAT ATTTATCAAT CACGGCGGGT TTCTAGAATG TATTGGCATT AAGCCTACAA	960
AACACACTCC TATAATATAC AAATATGACC TCAACCCGTG AATTCCAACA AAAAAACCAA	1020
CCCAACCAAA CCAAACTATT CCTCAAACAA CAGTGCTCAA TAGTTAAGAA GGAGCTAATC	1080
CATTITAGTA ATTAAAATA AAAGTAAAGC CAATAACATA AATTGGGGCA AATACAAAGA	1140
TGGCTCTTAG CAAAGTCAAG TTGAATGATA CATTAAATAA GGATCAGCTG CTGTCATCCA	1200
GCAAATACAC TATTCAACGT AGTACAGGAG ATAATATTGA CACTCCCAAT TATGATGTGC	1260
AAAAACACCT AAACAAACTA TGTGGTATGC TATTAATCAC TGAAGATGCA AATCATAAAT	1320
TCACAGGATT AATAGGTATG TTATATGCTA TGTCCAGGTT AGGAAGGGAA GACACTATAA	1380
AGATACTTAA AGATGCTGGA TATCATGTTA AAGCTAATGG AGTAGATATA ACAACATATC	1440
GTCAAGATAT AAATGGAAAG GAAATGAAAT TCGAAGTATT AACATTATCA AGCTTGACAT	1500
CAGAAATACA AGTCAATATT GAGATAGAAT CTAGAAAGTC CTACAAAAAA ATGCTAAAAG	1560
AGATGGGAGA AGTGGCTCCA GAATATAGGC ATGATTCTCC AGACTGTGGG ATGATAATAC	1620
TGTGTATAGC TGCACTTGTG ATAACCAAAT TAGCAGCAGG AGACAGATCA GGTCTTACAG	1680
CAGTAATTAG GAGGGCAAAC AATGTCTTAA AAAACGAAAT AAAACGATAC AAGGGCCTCA	1740
TACCAAAGGA TATAGCTAAC AGTTTTTATG AAGTGTTTGA AAAACACCCT CATCTTATAG	1800

- 348 -

ATGTTTTCGT GCACTTTGGC	: ATTGCACAAT	CATCCACAAG	AGGGGGTAGT	AGAGTTGAAG	1860
GAATCTTTGC AGGATTGTTT	ATGAATGCCT	ATGGTTCAGG	GCAAGTAATG	CTAAGATGGG	1920
GAGTTTTAGC CAAATCTGTA	AAAAATATCA	TGCTAGGACA	TGCTAGTGTC	CAGGCAGAAA	1980
TGGAGCAAGT TGTGGAAGTC	TATGAGTATG	CACAGAAGTT	GGGAGGAGAA	GCTGGATTCT	2040
ACCATATATT GAACAATCC	AAAGCATCAT	TGCTGTCATT	AACTCAATTT	CCCAACTTCT	2100
CAAGTGTGGT CCTAGGCAAT	GCAGCAGGTC	TAGGCATAAT	GGGAGAGTAT	AGAGGTACAC	2160
CAAGAAACCA GGATCTTTAT	GATGCAGCTA	AAGCATATGC	AGAGCAACTC	AAAGAAAATG	2220
GAGTAATAAA CTACAGTGTA	TTAGACTTAA	CAGCAGAAGA	ATTGGAAGCC	ATAAAGCATC	2280
AACTCAACCC CAAAGAAGAT	GATGTAGAGC	TTTAAGTTAA	CAAAAAATAC	GGGGCAAATA	2340
AGTCAACATG GAGAAGTTTG	CACCTGAATT	TCATGGAGAA	GATGCAAATA	ACAAAGCTAC	2400
CAAATTCCTA GAATCAATAA	AGGGCAAGTT	CGCATCATCC	AAAGATCCTA	AGAAGAAAGA	2460
TAGCATAATA TCTGTTAACT	CAATAGATAT	AGAAGTAACT	AAAGAGAGCC	CGATAACATC	2520
TGGCACCAAC ATCATCAATC	CAACAAGTGA	AGCCGACAGT	ACCCCAGAAA	CAAAAGCCAA	2580
CTACCCAAGA AAACCCCTAG	TAAGCTTCAA	AGAAGATCTC	ACCCCAAGTG	ACAACCCTTT	2640
TTCTAAGTTG TACAAGGAAA	CAATAGAAAC	ATTTGATAAC	AATGAAGAAG	AATCTAGCTA	2700
CTCATATGAA GAGATAAATG	ATCAAACAAA	TGACAACATT	ACAGCAAGAC	TAGATAGAAT	2760
TGATGAAAAA TTAAGTGAAA	TATTAGGAAT	GCTCCATACA	TTAGTAGTTG	CAAGTGCAGG	2820
ACCCACTTCA GCTCGCGATG	GAATAAGAGA	TGCTATGGTT	GGTCTAAGAG	AAGAGATGAT	2880
AGAAAAATA AGAGCGGAAG	CATTAATGAC	CAATGATAGG	TTAGAGGCTA	TGGCAAGACT	2940
TAGGAATGAG GAAAGCGAAA	AAATGGCAAA	AGACACCTCA	GATGAAGTGT	CTCTTAATCC	3000
AACTTCCAAA AAATTGAGTG	ACTTGTTGGA	AGACAACGAT	AGTGACAATG	ATCTATCACT	3060
TGATGATTTT TGATCAGCGA	. TCAACTCACT	CAGCAATCAA	CAACATCAAT	AAAACAGACA	3120
TCAATCCATT GAATCAACTG	CCAGACCGAA	CAAACAAACG	TCCATCAGTA	GAACCACCAA	3180
CCAATCAATC AACCAATTGA	TCAATCAGCA	ACCCGACAAA	ATTAACAATA	TAGTAACAAA	3240
AAAAGAACAA GATGGGGCAA	ATATGGAAAC	ATACGTGAAC	AAGCTTCACG	AAGGCTCCAC	3300
ATACACAGCA GCTGTTCAGT	ACAATGTTCT	AGAAAAAGAT	GATGATCCTG	CATCACTAAC	3360

- 349 -

AATATGGGTG	CCTATGTTCC	AGTCATCTG	GCCAGCAGAC	TTGCTCATAA	AAGAACTTGC	3420
AAGCATCAAT	TACTAGTGA	AGCAGATCTC	TACGCCCAAA	GGACCTTCAC	TACGAGTCAC	3480
GATTAACTCA	AGAAGTGCTG	TGCTGGCTCA	AATGCCTAGT	AATTTCATCA	TAAGCGCAAA	3540
TGTATCATTA	GATGAAAGAA	GCAAATTAGO	ATATGATGTA	ACTACACCTT	GTGAAATCAA	3600
AGCATGCAGT	CTAACATGCT	TAAAAGTAAA	AAGTATGTTA	ACTACAGTCA	AAGATCTTAC	3660
CATGAAGACA	TTCAACCCCA	CTCATGAGAT	CATTGCTCTA	TGTGAATTTG	AAAATATTAT	3720
GACATCAAAA	AGAGTAATAA	TACCAACCTA	TCTAAGATCA	ATTAGTGTCA	AGAACAAGGA	3780
TCTGAACTCA	CTAGAAAATA	TAGCAACCAC	CGAATTCAAA	AATGCTATCA	CCAATGCAAA	3840
AATTATTCCT	TATGCAGGAT	TAGTGTTAGT	TATCACAGTT	ACTGACAATA	AAGGAGCATT	3900
CAAATATATC	AAACCACAGA	GTCAATTTAT	AGTAGATCTT	GGTGCCTACC	TAGAAAAAGA	3960
GAGCATATAT	TATGTGACTA	CTAATTGGAA	GCATACAGCT	ACACGTTTTT	CAATCAAACC	4020
ACTAGAGGAT	TAAACTTAAT	TATCAACACT	GAATGACAGG	TCCACATATA	TCCTCAAACT	4080
ACACACTATA	TCCAAACATC	ATAAACATCT	ACACTACACA	CTTCATCACA	CAAACCAATC	4140
CCACTCAAAA	TCCAAAATCA	CTACCAGCCA	CTATCTGCTA	GACCTAGAGT	GCGAATAGGT	4200
AAATAAAACC	AAAATATGGG	GTAAATAGAC	ATTAGTTAGA	GTTCAATCAA	TCTTAACAAC	4260
CATTTATACC	GCCAATTCAA	CACATATACT	ATAAATCTTA	AAATGGGAAA	TACATCCATC	4320
			TATTTTACAC			4380
CTAATCTTTT	TACTAATTAT	AATCACTATT	ATGATTGCAA	TACTAAATAA	GCTAAGTGAA	4440
			CTAGGACAGA			4500
			ATCCTGTATA			4560
			ACGCTAACTA			4620
					CATTGGGGCA	4680
					AAAGACCTGG	4740
					TTTAAAATCT	4800
			ATGATAATCT			4860
GCCATAATAT	TCATCATCTC	TGCCAATCAC	AAAGTTACAC	TAACAACGGT	CACAGTTCAA	4920

- 350 -

ACAATAAAAA	ACCACACTGA	AAAAAACATC	ACCACCTACC	CTACTCAAGT	CTCACCAGAA	4980
AGGGTTAGTT	CATCCAAGCA	ACCCACAACC	ACATCACCAA	TCCACACAAG	TTCAGCTACA	5040
ACATCACCCA	ATACAAAATC	AGAAACACAC	CATACAACAG	CACAAACCAA	AGGCAGAACC	5100
ACCACTTCAA	CACAGACCAA	CAAGCCAAGC	ACAAAACCAC	GTCCAAAAAA	TCCACCAAAA	5160
AAAGATGATT	ACCATTTTGA	AGTGTTCAAC	TTCGTTCCCT	GCAGTATATG	TGGCAACAAT	5220
CAACTTTGCA	AATCCATCTG	CAAAACAATA	CCAAGCAACA	AACCAAAGAA	GAAACCAACC	5280
ATCAAACCCA	CAAACAAACC	AACCACCAAA	ACCACAAACA	AAAGAGACCC	AAAAACACCA	5340
GCCAAAACGA	CGAAAAAAGA	AACTACCACC	AACCCAACAA	AAAAACTAAC	CCTCAAGACC	5400
ACAGAAAGAG	ACACCAGCAC	CTCACAATCC	ACTGCACTCG	ACACAACCAC	ATTAAAACAC	5460
ACAGTCCAAC	AGCAATCCCT	CCTCTCAACC	ACCCCCGAAA	ACACACCCAA	CTCCACACAA	5520
ACACCCACAG	CATCCGAGCC	CTCCACACCA	AACTCCACCC	AAAAAACCCA	GCCACATGCT	5580
TAGTTATTCA	AAAACTACAT	CTTAGCAGAG	AACCGTGATC	TATCAAGCAA	GAACGAAATT	5640
AAACCTGGGG	CAAATAACCA	TGGAGTTGAT	GATCCACAAG	TCAAGTGCAA	TCTTCCTAAC	5700
TCTTGCTATT	AATGCATTGT	ACCTCACCTC	AAGTCAGAAC	ATAACTGAGG	AGTTTTACCA	5760
ATCGACATGT	AGTGCAGTTA	GCAGAGGTTA	TTTTAGTGCT	TTAAGAACAG	GTTGGTATAC	5820
TAGTGTCATA	ACAATAGAAT	TAAGTAATAT	AAAAGAAACC	AAATGCAATG	GAACTGACAC	5880
TAAAGTAAAA	CTTATGAAAC	AAGAATTAGA	TAAGTATAAG	AATGCAGTAA	CAGAATTACA	5940
GCTACTTATG	CAAAACACAC	CAGCTGTCAA	CAACCGGGCC	AGAAGAGAAG	CACCACAGTA	6000
TATGAACTAC	ACAATCAATA	CCACTAAAA	CCTAAATGTA	TCAATAAGCA	AGAAGAGGAA	6060
ACGAAGATTT	CTAGGCTTCT	TGTTAGGTGT	GGGATCTGCA	ATAGCAAGTG	GTATAGCTGT	6120
ATCAAAAGTT	CTACACCTTG	AAGGAGAAGT	GAACAAGATC	AAAAATGCTT	TGTTGTCTAC	6180
AAACAAAGCT	GTAGTCAGTT	TATCAAATGG	GGTCAGTGTT	TTAACCAGCA	AAGTGTTAGA	6240
TCTCAAGAAT	TACATAAATA	ACCAATTATT	ACCCATAGTA	AATCAACAGA	GCTGTCGCAT	6300
CTCCAACATT	GAAACAGTTA	TAGAATTCCA	GCAGAAGAAC	AGCAGATTGT	TGGAAATCAC	6360
CAGAGAATTT	AGTGTCAATG	CAGGTGTAAC	AACACCTTTA	AGCACTTACA	TGTTGACAAA	6420
CAGTGAGTTA	CTATCATTAA	TCAATGATAT	GCCTATAACA	AATGATCAGA	TAATTAAAA	6480

- 351 -

GTCAAGCAAT GTTCAGATAG TAAGGCAACA AAGTTATTCC ATCATGTCTA TAATAAAGGA	6540
AGAAGTCCTT GCATATGTTG TACAGCTGCC TATCTATGGT GTAATAGATA CACCTTGCTG	6600
GAAATTGCAC ACATCGCCTC TATGCACTAC CAACATCAAA GAAGGATCAA ATATTTGTTT	6660
AACAAGGACT GATAGAGGAT GGTATTGTGA TAATGCAGGA TCAGTATCCT TCTTTCCACA	6720
GGCTGACACT TGTAAAGTAC AGTCCAATCG AGTATTTTGT GACACTATGA ACAGTTTGAC	6780 ·
ATTACCAAGT GAAGTCAGCC TTTGTAACAC TGACATATTC AATTCCAAGT ATGACTGCAA	6840
ANTINIGACA TCAAAAACAG ACATAAGCAG CTCAGTAATI ACTICICITG GAGCTATAGI	6900
GTCATGCTAT GGTAAAACTA AATGCACTGC ATCCAACAAA AATCGTGGGA TTATAAAGAC	6960
ATTTTCTAAT GGTTGTGACT ATGTGTCAAA CAAAGGAGTA GATACTGTGT CAGTGGGCAA	7020
CACTITATAC TATGTAAACA AGCTGGAAGG CAAGAACCTT TATGTAAAAG GGGAACCTAT	7080
AATAAATTAC TATGACCCTC TAGTGTTTCC TTCTGATGAG TTTGATGCAT CAATATCTCA	7140
AGTCAATGAA AAAATCAATC AAAGTTTAGC TTTTATTCGT AGATCTGATG AATTACTACA	7200
TAATGTAAAT ACTGGCAAAT CTACTACAAA TATTATGATA ACTACAATTA TTATAGTAAT	7260
CATTGTAGTA TIGTTATCAT TAATAGCTAT TGGTTTACTG TTGTATTGTA	7320
CACACCAGTT ACACTAAGCA AAGACCAACT AAGTGGAATC AATAATATTG CATTCAGCAA	7380
ATAGACAAAA AACCACCTGA TCATGTTTCA ACAACAATCT GCTGACCACC AATCCCAAAT	7440
CAACTTACAA CAAATATTTC AACATCACAG TACAGGCTGA ATCATTTCCT CACATCATGC	7500
TACCCACATA ACTAAGCTAG ATCCTTAACT TATAGTTACA TAAAAACCTC AAGTATCACA	7560
ATCAACCACT AAATCAACAC ATCATTCACA AAATTAACAG CTGGGGCAAA TATGTCGCGA	7620
AGAAATCCTT GTAAATTTGA GATTAGAGGT CATTGCTTGA ATGGTAGAAG ATGTCACTAC	7680
AGTCATAATT ACTTTGAATG GCCTCCTCAT GCATTACTAG TGAGGCAAAA CTTCATGTTA	7740
AACAAGATAC TCAAGTCAAT GGACAAAAGC ATAGACACTT TGTCTGAAAT AAGTGGAGCT	7800
GCTGAACTGG ATAGAACAGA AGAATATGCT CTTGGTATAG TTGGAGTGCT AGAGAGTTAC	7860
ATAGGATCTA TAAACAACAT AACAAAACAA TCAGCATGTG TTGCTATGAG TAAACTTCTT	7920
ATTGAGATCA ATAGTGATGA CATTAAAAAG CTTAGAGATA ATGAAGAACC CAATTCACCT	7980
AAGATAAGAG TGTACAATAC TGTTATATCA TACATTGAGA GCAATAGAAA AAACAACAAG	8040

- 352 -

CAAACCATCC	ATCTGCTCAA	GAGACTACCA	GCAGACGTGC	TGAAGAAGAC	AATAAAGAAC	8100
ACATTAGATA	TCCACAAAAG	CATAACCATA	AGCAATCCAA	AAGAGTCAAC	TGTGAATGAT	8160
CAAAATGACC	AAACCAAAAA	TAATGATATT	ACCGGATAAA	TATCCTTGTA	GTATATCATC	8220
CATATTGATC	TCAAGTGAAA	GCATGGTTGC	TACATTCAAT	CATAAAAACA	TATTACAATT	8280
TAACCATAAC	TATTTGGATA	ACCACCAGCG	TTTATTAAAT	CATATATTTG	ATGAAATTCA	8340
TTGGACACCT	AAAAACTTAT	TAGATGCCAC	TCAACAATTT	CTCCAACATC	TTAACATCCC	8400
TGAAGATATA	TATACAGTAT	ATATATTAGT	GTCATAATGC	TTGACCATAA	CGACTCTATG	8460
TCATCCAACC	ATAAAACTAT	TTTGATAAGG	TTATGGGACA	AAATGGATCC	CATTATTAAT	8520
GGAAACTCTG	CTAATGTGTA	TCTAACTGAT	AGTTATTTAA	AAGGTGTTAT	CTCTTTTTCA	8580
GAGTGTAATG	CTTTAGGGAG	TTATCTTTTT	AACGGCCCTT	ATCTTAAAAA	TGATTACACC	8640
AACTTAATTA	GTAGACAAAG	CCCACTACTA	GAGCATATGA	ATCTTAAAAA	ACTAACTATA	8700
ACACAGTCAT	TAATATCTAG	ATATCATAAA	GGTGAACTGA	AATTAGAAGA	ACCAACTTAT	8760
TTCCAGTCAT	TACTTATGAC	ATATAAAAGT	ATGTCCTCGT	CTGAACAAAT	TGCTACAACT	8820
AACTTACTTA	AAAAAATAAT	ACGAAGAGCC	ATAGAAATAA	GTGATGTAAA	GGTGTACGCC	8880
ATCTTGAATA	AACTAGGATT	AAAGGAAAAG	GACAGAGTTA	AGCCCAACAA	TAATTCAGGT	8940
GATGAAAACT	CAGTACTTAC	AACTATAATT	AAAGATGATA	TACTTTCGGC	TGTGGAAAAC	9000
AATCAATCAT	ATACAAATTC	AGACAAAAGT	CACTCAGTAA	ATCAAAATAT	CACTATCAAA	9060
ACAACACTCT	TGAAAAAATT	GATGTGTTCA	ATGCAACATC	CTCCATCATG	GTTAATACAC	9120
TGGTTCAATT	TATATACAAA	ATTAAATAAC	ATATTAACAC	AATATCGATC	AAATGAGGTA	9180
AAAAGTCATG	GGTTTATATT	AATAGATAAT	CAAACTTTAA	GTGGTTTTCA	GTTTATTTTA	9240
AATCAATATG	GTTGTATCGT	TTATCATAAA	GGACTCAAAA	AAATCACAAC	TACTACTTAC	9300
AATCAATTTT	TGACATGGAA	AGACATCAGC	CTTAGCAGAT	TAAATGTTTG	CTTAATTACT	9360
TGGATAAGTA	ATTGTTTAAA	TACATTAAAC	AAAAGCTTAG	GGCTGAGATG	TGGATTCAAT	9420
					ATTTCATAAT	
					TCTAAACATA	9540
ACAGAAGAAG	ATCAATTTAG	GAAACGATTT	TATAATAGCA	TGCTAAATAA	CATCACAGAT	9600

- 353 -

GCAGCTATTA	AGGCTCAAAA	GGACCTACTA	TCAAGAGTAT	GTCACACTTT	ATTAGACAAG	9660
ACAGTGTCTG	ATAATATCAT	AAATGGTAAA	TGGATAATCC	TATTAAGTAA	ATTTCTTAAA	9720
TTGATTAAGC	TTGCAGGTGA	TAATAATCTC	AATAACTTGA	GTGAGCTATA	TTTTCTCTTC	9780
AGAATCTTTG	GACATCCAAT	GGTCGATGAA	AGACAAGCAA	TGGATTCTGT	AAGAATTAAC	9840
TGTAATGAAA	CTAAGTTCTA	CTTATTAAGT	AGTCTAAGTA	CATTAAGAGG	TGCTTTCATT	9900
TATAGAATCA	TAAAAGGGTT	TGTAAATACC	TACAACAGAT	GGCCCACCTT	AAGGAATGCT	9960
ATTGTCCTAC	CTCTAAGATG	GTTAAACTAC	TATAAACTTA	ATACTTATCC	ATCTCTACTT	10020
GAAATCACAG	AAAATGATTT	GATTATTTTA	TCAGGATTGC	GGTTCTATCG	TGAGTTTCAT	10080
CTGCCTAAAA	AAGTGGATCT	TGAAATGATA	ATAAATGACA	AAGCCATTTC	ACCTCCAAAA	10140
GATCTAATAT	GGACTAGTTT	TCCTAGAAAT	TACATGCCAT	CACATATACA	AAATTATATA	10200
GAACATGAAA	AGTTGAAGTT	CTCTGAAAGC	GACAGATCGA	GAAGAGTACT	AGAGTATTAC	10260
TTGAGAGATA	ATAAATTCAA	TGAATGCGAT	CTATACAATT	GTGTAGTCAA	TCAAAGCTAT	10320
CTCAACAACT	CTAATCACGT	GGTATCACTA	ACTGGTAAAG	AAAGAGAGCT	CAGTGTAGGT	10380
AGAATGTTTG	CTATGCAACC	AGGTATGTTT	AGGCAAATCC	AAATCTTAGC	AGAGAAAATG	10440
ATAGCTGAAA	ATATTTTACA	ATTCTTCCCT	GAGAGTTTGA	CAAGATATGG	TGATCTAGAG	10500
CTTCAAAAGA	TATTAGAATT	AAAAGCAGGA	ATAAGCAACA	AGTCAAATCG	TTATAATGAT	10560
AACTACAACA	ATTATATCAG	TAAATGTTCT	ATCATTACAG	ATCTTAGCAA	ATTCAATCAG	10620
GCATTTAGAT	ATGAAACATC	ATGTATCTGC	AGTGATGTAT	TAGATGAACT	GCATGGAGTA	10680
CAATCTCTGT	TCTCTTGGTT	GCATTTAACA	ATACCTCTTG	TCACAATAAT	ATGTACATAT	10740
AGACATGCAC	CTCCTTTCAT	AAAGGATCAT	GTTGTTAATC	TTAATGAGGT	TGATGAACAA	10800
AGTGGATTAT	ACAGATATCA	TATGGGTGGT	ATTGAGGGCT	GGTGTCAAAA	ACTGTGGACC	10860
ATTGAAGCTA	TATCATTATT	AGATCTAATA	TCTCTCAAAG	GGAAATTCTC	TATCACAGCT	10920
CTGATAAATG	GTGATAATCA	GTCAATTGAT	ATAAGCAAAC	CAGTTAGACT	TATAGAGGGT	10980
CAGACCCATG	CACAAGCAGA	TTATTTGTTA	GCATTAAATA	GCCTTAAATT	GTTATATAAA	11040
GAGTATGCAG	GTATAGGCCA	TAAGCTTAAG	GGAACAGAGA	CCTATATATC	CCGAGATATG	11100
CAGTTCATGA	GCAAAACAAT	CCAGCACAAT	GGAGTGTACT	ATCCAGCCAG	TATCAAAAAA	11160

- 354 -

GTCCTGAGAG	TAGGTCCATG	GATAAACACG	ATACTTGATG	ATTTTAAAGT	TAGTTTAGAA	11220
TCTATAGGCA	GCTTAACACA	GGAGTTAGAA	TACAGAGGAG	AAAGCTTATT	ATGCAGTTTA	11280
ATATTTAGGA	ACATTTGGTT	ATACAATCAA	ATTGCTTTGC	AACTCCGAAA	TCATGCATTA	11340
TGTAACAATA	AGCTATATTT	AGATATATTG	AAAGTATTAA	AACACTTAAA	AACTTTTTT	11400
AATCTTGATA	GCATTGATAT	GGCTTTATCA	TTGTATATGA	ATTTGCCTAT	GCTGTTTGGT	11460
GGTGGTGATC	CTAATTTGTT	ATATCGAAGC	TTTTATAGGA	GAACTCCAGA	CTTCCTTACA	11520
GAAGCTATAG	TACATTCAGT	GTTTGTGTTG	AGCTATTATA	CTGGTCACGA	TTTACAAGAT	11580
AAGCTCCAGG	ATCTTCCAGA	TGATAGACTG	AACAAATTCT	TGACATGTGT	CATCACATTT	11640
GATAAAAATC	CCAATGCCGA	GTTTGTAACA	TTGATGAGGG	ATCCACAGGC	TTTAGGGTCT	11700
GAAAGGCAAG	CTAAAATTAC	TAGTGAGATT	AATAGATTAG	CAGTAACAGA	AGTCTTAAGT	11760
ATAGCCCCAA	ACAAAATATT	TTCTAAAAGT	GCACAACATT	ATACTACCAC	TGAGATTGAT	11820
CTAAATGACA	TTATGCAAAA	TATAGAACCA	ACTTACCCTC	ATGGATTAAG	AGTTGTTTAT	11880
GAAAGTTTAC	CTTTTTATAA	AGCAGAAAAA	ATAGTTAATC	TTATATCAGG	AACAAAATCC	11940
ATAACTAATA	TACTTGAAAA	AACATCAGCA	ATAGATACAA	CTGATATTAA	TAGGGCTACT	12000
GATATGATGA	GGAAAAATAT	AACTTTACTT	ATAAGGATAC	TTCCACTAGA	TTGTAACAAA.	12060
GACAAAAGAG	AGTTATTAAG	TTTAGAAAAT	CTTAGTATAA	CTGAATTAAG	CAAGTATGTA.	12120
AGAGAAAGAT	CTTGGTCATT	ATCCAATATA	GTAGGAGTAA	CATCGCCAAG	TATTATGTTC	12180
ACAATGGACA	TTAAATATAC	AACTAGCACT	ATAGCCAGTG	GTATAATAAT	AGAAAATAT	12240
AATGTTAATA	GTTTAACTCG	TGGTGAAAGA	GGACCCACCA	AGCCATGGGT	AGGCTCATCC	12300
ACGCAGGAGA	AAAAAACAAT	GCCAGTGTAC	AACAGACAAG	TTTTAACCAA	AAAGCAAAGA	12360
GACCAAATAG	ATTTATTAGC	AAAATTAGAC	TGGGTATATG	CATCCATAGA	CAACAAAGAT	12420
GAATTCATGG	AAGAACTGAG	TACTGGAACA	CTTGGACTGT	CATATGAAAA	AGCCAAAAAG	12480
TTGTTTCCAC	AATATCTAAG	TGTCAATTAT	TTACACCGTT	TAACAGTCAG	TAGTAGACCA	12540
TGTGAATTCC	CTGCATCAAT	ACCAGCTTAT	AGAACAACAA	ATTATCATTT	TGATACTAGT	12600
CCTATCAATC	ATGTATTAAC	AGAAAGTAT	GGAGATGAAG	ATATCGACAT	TGTGTTTCAA.	12660
AATTGCATAA	GTTTTGGTCT	TAGCCTGATG	TCGGTTGTGG	AACAATTCAC	AAACATATGT	12720

- 355 -

CCTAATAGA	A TTATTCTCA	r accgaaget	G AATGAGATA	C ATTTGATGA	A ACCTCCTATA	12780
TTTACAGGAG	ATGTTGATA	CATCAAGTTO	G AAGCAAGTGI	A TACAAAAGC	A GCACATGTTC	12840
CTACCAGATA	AAATAAGTT	r aacccaata	GTAGAATTA	TCTTAAGTA	A CAAAGCACTT	12900
AAATCTGGAT	CTCACATCAL	A CTCTAATTT	A ATATTAGTA	ATAAAATGTO	C TGATTATTTT	12960
CATAATGCTT	AATTTTAAC	TACTAATTT	A GCTGGACATI	GGATTCTGAT	TATTCAACTT	13020
ATGAAAGATT	CAAAAGGTAT	TTTTGAAAA	A GATTGGGGA	G AGGGGTACAT	F AACTGATCAT	13080
ATGTTCATTA	ATTTGAATG	TTTCTTTAA1	gcttataaga	CTTATTTGCT	ATGTTTTCAT	13140
AAAGGTTATG	GTAAAGCAA	ATTAGAATG1	GATATGAACA	CTTCAGATCT	CTCTTTGTGTT	13200
TTGGAGTTAA	TAGACAGTAG	CTACTGGAAA	TCTATGTCTA	AAGTTTTCCT	AGAACAAAA	13260
GTCATAAAAT	ACATAGTCAA	TCAAGACACA	AGTTTGCGTA	GAATAAAAGG	CTGTCACAGT	13320
TTTAAGTTGT	GGTTTTTAAA	ACGCCTTAAT	AATGCTAAAT	' TTACCGTATG	CCCTTGGGTT	13380
GTTAACATAG	ATTATCACCO	AACACACATG	AAAGCTATAT	TATCTTACAT	AGATTTAGTT	13440
AGAATGGGGT	TAATAAATGI	AGATAAATTA	ACCATTAAAA	ATARARACAA	ATTCAATGAT	13500
GAATTTTACA	CATCAAATCT	CTTTTACATT	AGTTATAACT	TTTCAGACAA	CACTCATTTG	13560
CTAACAAAAC	AAATAAGAAT	TGCTAATTCA	GAATTAGAAG	ATAATTATAA	CAAACTATAT	13620
CACCCAACCC	CAGAAACTTT	AGAAAATATG	TCATTAATTC	CTGTTAAAAG	TAATAATAGT	13680
AACAAACCTA	AATTTTGTAT	AAGTGGAAAT	ACCGAATCTA	TGATGATGTC	AACATTCTCT	13740
AGTAAAATGC	ATATTAAATC	TTCCACTGTT	ACCACAAGAT	TCAATTATAG	CAAACAAGAC	13800
		TGTTGTGATA				13860
GCAAAATCTA	ACCAACTTTA	CACCACCACT	TCACATCAGA	CATCTTTAGT	AAGGAATAGT	13920
		TCCTTGGCAT				13980
					TAAGGACCCC	14040
					GGTAGTAGAA	14100
					TCATAGTTTA	14160
					TGAGAATTTA	14220
ACCATTCCTG	CTACAGATGC	AACTAATAAC	ATTCATTGGT	CTTATTTACA	TATAAAATTT	14280

- 356 -

GCAGAACCTA	TTAGCATCTT	TGTCTGCGAT	GCTGAATTAC	CTGTTACAGC	CAATTGGAGT	14340
AAAATTATAA	TTGAATGGAG	TAAGCATGTA	AGAAAGTGCA	AGTACTGTTC	TTCTGTAAAT	14400
AGATGCATTT	TAATTGCAAA	ATATCATGCT	CAAGATGACA	TTGATTTCAA	ATTAGATAAC	14460
ATTACTATAT	TAAAAACTTA	CGTGTGCCTA	GGTAGCAAGT	TAAAAGGATC	TGAAGTTTAC	14520
TTAATCCTTA	CAATAGGCCC	TGCAAATATA	CTTCCTGTTT	TTGATGTTGT	ACAAAATGCT	14580
AAATTGATAC	TTTCAAGAAC	TAAAAATTTC	ATTATGCCTA	AAAAAACTGA	CAAGGAATCT	14640
ATCGATGCAG	ATATTAAAAG	CTTAATACCT	TTCCTTTGTT	ACCCTATAAC	AAAAAAAGGA	14700
ATTAAGACTT	CATTGTCAAA	ATTGAAGAGT	GTAGTTAATG	GAGATATATT	ATCATATTCT	14760
ATAGCTGGAC	GTAATGAAGT	ATTCAGCAAC	AAGCTTATAA	ACCACAAGCA	TATGAATATC	14820
CTAAAATGGC	TAGATCATGT	TTTAAATTTT	AGATCAGCTG	AACTTAATTA	CAATCATTTA	14880
TACATGATAG	AGTCCACATA	TCCTTACTTA	AGTGAATTGT	TAAATAGTTT	AACAACCAAT	14940
GAGCTCAAGA	AGCTGATTAA	AATAACAGGT	AGTGTGCTAT	ACAACCTTCC	CAACGAACAG	15000
TAGTTTAAAA	TATCATTAAC	AAGTTTGGTC	AAATTTAGAT	GCTAACACAT	CATTATATTA	15060
TAGTTATTAA	AAAATATACA	AACTTTTCAA	TAATTTAGCA	TATTGATTCC	AAAATTATCA	15120
TTTTAGTCTT	AAGGGGTTAA	ATAAAAGTCT	AAAACTAACA	ATTATACATG	TGCATTCACA	15180
ACACAACGAG	ACATTAGTTT	TTGACACTTT	TTTTCTCGT			15219

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15

Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly

- 357 -

			20					25					30		
Sex	Туз	35	ı Phe	a Asr	a Gly	Pro	40	Let	Ly:	a Asr	Asr	45	Thi	r Ası	ı Leu
Ile	Ser 50	Arg	g Glr	ı Ser	Pro	Leu 55	. Leu	Glu	His	Met	Asn 60	Lev	Lys	Lys	3 Leu
Thr 65	Ile	Thi	Glr	Ser	Leu 70	Ile	Ser	Arg	Туг	75	Lys	Gly	Glu	Leu	Lys 80
Leu	Glu	Glu	Pro	Thr 85	Tyr	Phe	Gln	Ser	Leu 90	Leu	Met	Thr	Туг	Lys 95	Ser
Met	Ser	Ser	Ser 100	Glu	Gln	Ile	Ala	Thr 105		Asn	Leu	Leu	Lys 110		Ile
Ile	Arg	Arg		Ile	Glu	Ile	Ser 120		Val	Lys	Val	Tyr 125		Ile	Leu
Asn	Lув 130	Leu	Gly	Leu	Lys	Glu 135	Lys	Asp	Arg	Val	Lys 140	Pro	Asn	Asn	Asn
Ser 145	Gly	Авр	Glu	Asn	Ser 150	Val	Leu	Thr	Thr	Ile 155	Ile	Lys	Asp	Asp	Ile 160
Leu	Ser	Ala	Val	Glu 165	Asn	Asn	Gln	Ser	Tyr 170	Thr	Asn	Ser	Asp	Lys 175	Ser
His	Ser	Val	Asn 180	Gln	Asn	Ile	Thr	Ile 185	Lys	Thr	Thr	Leu	Leu 190	Lys	Lys
Leu	Met	Сув 195	Ser	Met	Gln	His	Pro 200	Pro	Ser	Trp	Leu	Ile 205	His	Trp	Phe
Asn	Leu 210	Tyr	Thr	Lys	Leu	Asn 215	Asn	Ile	Leu	Thr	Gln 220	Tyr	Arg	Ser	Asn
Glu 225	Val	Lys	Ser	His	Gly 230	Phe	Ile	Leu	Ile	Asp 235	Asn	Gln	Thr	Leu	Ser 240
Gly	Phe	Gln	Phe	Ile 245	Leu	Asn	Gln	Tyr	Gly 250	Сув	Ile	Val	Tyr	His 255	Lys
Gly	Leu	Lys	Lys 260	Ile	Thr	Thr	Thr	Thr 265	Tyr	Asn	Gln	Phe	Leu 270	Thr	Trp
Lys	Asp	Ile 275	Ser	Leu	Ser	Arg	Leu 280	Asn	Val	Сув		Ile 285	Thr	Trp	Ile
Ser	Asn 290	Сув	Leu	Asn	Thr	Leu 295	Asn	Lys	Ser		Gly 300	Leu	Arg	Сув	Gly

- 358 -

Phe 305	Asn	Asn	Val	Val	Leu 310	Ser	Gln	Leu	Phe	Leu 315	Tyr	Gly	Asp	Сув	11e 320
Leu	Lys	Leu	Phe	His 325	Asn	Glu	Gly	Phe	Tyr 330	Ile	Ile	Lys	Glu	Val 335	Glu
Gly	Phe	Ile	Met 340	Ser	Leu	Ile	Leu	Asn 345	Ile	Thr	Glu	Glu	<b>Asp</b> 350	Gln	Phe
Arg	Lув	Arg 355	Phe	Tyr	Asn	Ser	Met 360	Leu	Asn	Asn	Ile	Thr 365	Двр	Ala	Ala
Ile	Lys 370	Ala	Gln	Lys	Asp	Leu 375	Leu	Ser	Arg	Val	Сув 380	His	Thr	Leu	Leu
Asp 385	Lys	Thr	Val	Ser	Двр 390	Asn	Ile	Ile	Asn	Gly 395	ГÀв	Trp	Ile	Ile	Leu 400
Leu	Ser	Lys	Phe	Leu 405	Lув	Leu	Ile	Lys	Leu 410	Ala	Gly	Авр	Asn	Asn 415	Leu
Asn	Asn	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
Met	Val	Asp 435	Glu	Arg	Gln	Ala	Met 440	Asp	Ser	Val	Arg	11e 445	Asn	Сув	Asn
Glu	Thr 450	Lys	Phe	Tyr	Leu	Leu 455	Ser	Ser	Leu	Ser	Thr 460	Leu	Arg	Gly	Ala
Phe 465	Ile	Tyr	Arg	Ile	11e 470	Lys	Gly	Phe	Val	Asn 475	Thr	Tyr	Asn	Arg	Trp 480
Pro	Thr	Leu	Arg	Asn 485	Ala	Ile	Val	Leu	Pro 490	Leu	Arg	Trp	Leu	Asn 495	Tyr
Tyr	Lys	Leu	Asn 500	Thr	Tyr	Pro	Ser	Leu 505	Leu	Glu	Ile	Thr	Glu 510	Asn	yab
Leu	Ile	Ile 515		Ser	Gly	Leu	Arg 520		Tyr	Arg	Glu	Phe 525	His	Leu	Pro
Lys	Lys 530		Asp	Leu	Glu	Met 535		Ile	Asn	Asp	<b>Lys</b> 540	Ala	Ile	Ser	Pro
Pro 545	_	Asp	Leu	Ile	Trp 550		Ser	Phe	Pro	Arg 555		Tyr	Met	Pro	Ser 560
His	Ile	Gln	Asn	Tyr 565		Glu	His	Glu	<b>Lys</b> 570		Lys	Phe	Ser	Glu 575	Ser

- 359 -

Asp	Arg	Ser	Arg 580		Val	Leu	Glu	585		Leu	Arg	J Yeb	590	-	3 Phe
Asn	Glu	Cys 595	Asp	Leu	Tyr	Asn	Сув 600		. Val	. Asn	Glr	Ser 605	-	: Leu	ABI
Asn	Ser 610		His	Val	Val	Ser 615		Thr	Gly	Lya	620		Glu	ı Lev	Sez
Val 625		Arg	Met	Phe	Ala 630	Met	Gln	Pro	Gly	Met 635		Arg	Gl:	Ile	640
			Glu	645					650					655	
			Thr 660					665					670		
		675	Gly				680					685			_
	690		Ile			695					700			_	
705			Phe		710					715					720
			His	725					730					735	
			Val 740					745					750		
		755	His				760					765			
	770		Tyr			775					780				
785			Glu		790					795					800
			Ile	805					810					815	
			Pro 820					825					830		
		835	Leu Gly				840					845			
	y	T16	GTA	TITE	n y B	neu	пЛв	GTA	TIII	GIU	THE	LAL	TTG	Ser	Arg

- 360 -

850		855	860		
Asp Met Gln 865	Phe Met Ser 870	Lys Thr Ile	Gln His Asn 875	Gly Val	Tyr Tyr 880
Pro Ala Ser	Ile Lys Lys 885	Val Leu Arg	Val Gly Pro 890	-	Asn Thr 895
Ile Leu Asp	Asp Phe Lys 900	Val Ser Leu 905	Glu Ser Ile	Gly Ser 910	Leu Thr
Gln Glu Leu 915	Glu Tyr Arg	Gly Glu Ser 920	Leu Leu Cys	Ser Leu 925	Ile Phe
Arg Asn Ile 930	Trp Leu Tyr	Asn Gln Ile 935	Ala Leu Gln 940	Leu Arg	Asn His
Ala Leu Cys 945	Asn Asn Lys 950	Leu Tyr Leu	Asp Ile Leu 955	Lys Val	Leu Lys 960
His Leu Lys	Thr Phe Phe 965	Asn Leu Asp	Ser Ile Asp 970		Leu Ser 975
Leu Tyr Met	Asn Leu Pro 980	Met Leu Phe 985	Gly Gly Gly	Asp Pro 990	Asn Leu
Leu Tyr Arg 995	Ser Phe Tyr	Arg Arg Thr 1000	Pro Asp Phe	Leu Thr 1005	Glu Ala
Ile Val His 1010	Ser Val Phe	Val Leu Ser 1015	Tyr Tyr Thr	-	Asp Leu
Gln Asp Lys 1025	Leu Gln Asp 103	Leu Pro Asp O	Asp Arg Leu 1035	Asn Lys	Phe Leu 1040
Thr Cys Val	Ile Thr Phe 1045	Asp Lys Asn	Pro Asn Ala 1050		Val Thr 1055
Leu Met Arg	Asp Pro Gln 1060	Ala Leu Gly 106	-	Gln Ala 1070	-
Thr Ser Glu 107	_	Leu Ala Val 1080	Thr Glu Val	Leu Ser 1085	Ile Ala
Pro Asn Lys 1090	Ile Phe Ser	Lys Ser Ala 1095	Gln His Tyr 1100		Thr Glu
Ile Asp Leu 1105	Asn Asp Ile 111	Met Gln Asn O	Ile Glu Pro 1115	Thr Tyr	Pro His 1120
Gly Leu Arg	Val Val Tyr 1125	Glu Ser Leu	Pro Phe Tyr 1130	-	Glu Lys 1135

- 361 -

- Ile Val Asn Leu Ile Ser Gly Thr Lys Ser Ile Thr Asn Ile Leu Glu 1140 1145 1150
- Lys Thr Ser Ala Ile Asp Thr Thr Asp Ile Asn Arg Ala Thr Asp Met 1155 1160 1165
- Met Arg Lys Asn Ile Thr Leu Leu Ile Arg Ile Leu Pro Leu Asp Cys 1170 1175 1180
- Asn Lys Asp Lys Arg Glu Leu Leu Ser Leu Glu Asn Leu Ser Ile Thr 1185 1190 1195 1200
- Glu Leu Ser Lys Tyr Val Arg Glu Arg Ser Trp Ser Leu Ser Asn Ile 1205 1210 1215
- Val Gly Val Thr Ser Pro Ser Ile Met Phe Thr Met Asp Ile Lys Tyr 1220 1225 1230
- Thr Thr Ser Thr Ile Ala Ser Gly Ile Ile Ile Glu Lys Tyr Asn Val 1235 1240 1245
- Asn Ser Leu Thr Arg Gly Glu Arg Gly Pro Thr Lys Pro Trp Val Gly 1250 1260
- Ser Ser Thr Gln Glu Lys Lys Thr Met Pro Val Tyr Asn Arg Gln Val 1265 1270 1275 1280
- Leu Thr Lys Lys Gln Arg Asp Gln Ile Asp Leu Leu Ala Lys Leu Asp 1285 1290 1295
- Trp Val Tyr Ala Ser Ile Asp Asn Lys Asp Glu Phe Met Glu Glu Leu 1300 1305 1310
- Ser Thr Gly Thr Leu Gly Leu Ser Tyr Glu Lys Ala Lys Lys Leu Phe 1315 1320 1325
- Pro Gln Tyr Leu Ser Val Asn Tyr Leu His Arg Leu Thr Val Ser Ser 1330 1335 1340
- Arg Pro Cys Glu Phe Pro Ala Ser Ile Pro Ala Tyr Arg Thr Thr Asn 1345 1350 1355 1360
- Tyr His Phe Asp Thr Ser Pro Ile Asn His Val Leu Thr Glu Lys Tyr 1365 1370 1375
- Gly Asp Glu Asp Ile Asp Ile Val Phe Gln Asn Cys Ile Ser Phe Gly 1380 1385 1390
- Leu Ser Leu Met Ser Val Val Glu Gln Phe Thr Asn Ile Cys Pro Asn 1395 1400 1405

- 362 -

Arg Ile Ile Leu Ile Pro Lys Leu Asn Glu Ile His Leu Met Lys Pro Pro Ile Phe Thr Gly Asp Val Asp Ile Ile Lys Leu Lys Gln Val Ile 1430 1435 Gln Lys Gln His Met Phe Leu Pro Asp Lys Ile Ser Leu Thr Gln Tyr 1445 1450 Val Glu Leu Phe Leu Ser Asn Lys Ala Leu Lys Ser Gly Ser His Ile Asn Ser Asn Leu Ile Leu Val His Lys Met Ser Asp Tyr Phe His Asn 1475 1480 Ala Tyr Ile Leu Ser Thr Asn Leu Ala Gly His Trp Ile Leu Ile Ile 1495 Gln Leu Met Lys Asp Ser Lys Gly Ile Phe Glu Lys Asp Trp Gly Glu 1505 1510 1515 1520 Gly Tyr Ile Thr Asp His Met Phe Ile Asn Leu Asn Val Phe Phe Asn 1525 Ala Tyr Lys Thr Tyr Leu Leu Cys Phe His Lys Gly Tyr Gly Lys Ala 1545 Lys Leu Glu Cys Asp Met Asn Thr Ser Asp Leu Leu Cys Val Leu Glu 1555 1560 Leu Ile Asp Ser Ser Tyr Trp Lys Ser Met Ser Lys Val Phe Leu Glu 1575 1580 Gln Lys Val Ile Lys Tyr Ile Val Asn Gln Asp Thr Ser Leu Arg Arg 1585 Ile Lys Gly Cys His Ser Phe Lys Leu Trp Phe Leu Lys Arg Leu Asn 1610 Asn Ala Lys Phe Thr Val Cys Pro Trp Val Val Asn Ile Asp Tyr His 1620 1625 1630 Pro Thr His Met Lys Ala Ile Leu Ser Tyr Ile Asp Leu Val Arg Met 1640 Gly Leu Ile Asn Val Asp Lys Leu Thr Ile Lys Asn Lys Asn Lys Phe 1650 1655 1660 Asn Asp Glu Phe Tyr Thr Ser Asn Leu Phe Tyr Ile Ser Tyr Asn Phe 1665 1670 1675

Ser Asp Asn Thr His Leu Leu Thr Lys Gln Ile Arg Ile Ala Asn Ser

- 363 -

	1.605			
	1685	1	690	1695
Glu Leu Glu Asp 170		Lys Leu T 1705	yr His Pro Thr	Pro Glu Thr 1710
Leu Glu Asn Met 1715	Ser Leu Ile	Pro Val Ly 1720	ys Ser Asn Asn 172	•
Pro Lys Phe Cys 1730	Ile Ser Gly 173		lu Ser Met Met 1740	Met Ser Thr
Phe Ser Ser Lys 1745	Met His Ile 1750	Lys Ser Se	or Thr Val Thr 1755	Thr Arg Phe 1760
Asn Tyr Ser Lys	Gln Asp Leu 1765		eu Phe Pro Ile 170	Val Val Ile 1775
Asp Lys Ile Ile 1780		Gly Asn Th	or Ala Lys Ser	Asn Gln Leu 1790
Tyr Thr Thr Thr 1795	Ser His Gln	Thr Ser Le	ou Val Arg Asn 180	
Leu Tyr Cys Met 1810	Leu Pro Trp 181		l Asn Arg Phe 1820	Asn Phe Val
Phe Ser Ser Thr 1825	Gly Cys Lys 1830	Ile Ser Il	e Glu Tyr Ile 1835	Leu Lys Asp 1840
Leu Lys Ile Lys	Asp Pro Ser 1845		a Phe Ile Gly 50	Glu Gly Ala 1855
Gly Asn Leu Leu 1860		Val Val Gl 1865	u Leu His Pro	Asp Ile Arg 1870
Tyr Ile Tyr Arg 1875	Ser Leu Lys	Авр Сув <b>А</b> в 1880	n Asp His Ser 1885	
Glu Phe Leu Arg	Leu Tyr Asn 1895		e Asn Ile Asp 1900	Tyr Gly Glu
Asn Leu Thr Ile : 1905	Pro Ala Thr 1910	Asp Ala Th	r Asn Asn Ile 1915	His Trp Ser 1920
Tyr Leu His Ile	Lys Phe Ala 1925	Glu Pro Ile 193		Val Cys Asp 1935
Ala Glu Leu Pro 1	Val Thr Ala	Asn Trp Ser 1945	Lys Ile Ile	Ile Glu Trp 1950
Ser Lys His Val	Arg Lys Cys	Lys Tyr Cys	S Ser Ser Val	Asn Arg Cys

1960

1965

1955

- 364 -

Ile Leu Ile Ala Lys Tyr His Ala Gln Asp Asp Ile Asp Phe Lys Leu 1970 1975 1980

Asp Asn Ile Thr Ile Leu Lys Thr Tyr Val Cys Leu Gly Ser Lys Leu 1985 1990 1995 2000

Lys Gly Ser Glu Val Tyr Leu Ile Leu Thr Ile Gly Pro Ala Asn Ile 2005 2010 2015

Leu Pro Val Phe Asp Val Val Gln Asn Ala Lys Leu Ile Leu Ser Arg 2020 2025 2030

Thr Lys Asn Phe Ile Met Pro Lys Lys Thr Asp Lys Glu Ser Ile Asp 2035 2040 2045

Ala Asp Ile Lys Ser Leu Ile Pro Phe Leu Cys Tyr Pro Ile Thr Lys 2050 2055 2060

Lys Gly Ile Lys Thr Ser Leu Ser Lys Leu Lys Ser Val Val Asn Gly 2065 2070 2075 2080

Asp Ile Leu Ser Tyr Ser Ile Ala Gly Arg Asn Glu Val Phe Ser Asn 2085 2090 2095

Lys Leu Ile Asn His Lys His Met Asn Ile Leu Lys Trp Leu Asp His 2100 2105 2110

Val Leu Asn Phe Arg Ser Ala Glu Leu Asn Tyr Asn His Leu Tyr Met 2115 2120 2125

Ile Glu Ser Thr Tyr Pro Tyr Leu Ser Glu Leu Leu Asn Ser Leu Thr 2130 2135 2140

Thr Asn Glu Leu Lys Lys Leu Ile Lys Ile Thr Gly Ser Val Leu Tyr 2145 2150 2155 2166

Asn Leu Pro Asn Glu Gln 2165

#### (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15219 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA (genomic)

- 365 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

60	AAATAAGAAC	AAAATGGGGC	ACATTCGAAA	ACAAACTTGC	AATGCGTACT	ACGGGAAAAA
120	TTCACTGAGC	TGGGGTGCAA	CAATCAGAAA	CTAACCTTTT	CTATTTAAGT	TTGATAAGTG
180	GTTAAAAATA	AAGTAGCATT	GACAATGACG	AAATTTATTT	TTAGATTACA	ATGATAAAGG
240	AGCAATACAT	TAGCCAAAGC	ACCAATGCAT	AATTCTTCTG	CTGATAAATT	ACATGTTATA
300	AGTGTGCCCT	CAAGCAGTGA	CATGTTATAA	AGTTTTATA	TAAACGGCAT	ACAATTAAAT
360	aaatggagga	CAATACTACA	ACAACAATGC	ATCTAACTTT	TTGTAGTGAA	GATAACAATA
420	AATGGATGAT	TAAACGGTTT	TGCTCTCAAT	GTTGACACAC	AATTGATTGA	TACATATGGG
480	TTATATGAAT	TAATGACTAA	AGTGACTCAG	TAAAAGACTA	TCAAATTTTC	aattgtgaaa
540	TCTAATTCAA	TTATGTTTAG	AATTCATGAA	GCTTGATCTC	ACTTACTTGG	CAAATATCTG
600	AAATGGGGCA	CATCAAAGGG	TATAAAAACT	TTTTAGTTAA	TTTATTACCA	TAGACATGTG
660	ACTATGCAAA	TGACAACACT	GCACTACAAA	CAAACCATGA	CCTAATCAAT	AATAAACTCA
720	TCTCTTACCA	AATAATAACA	CAATGGATTC	AGACCCCTGT	CACAGACATG	GATTGATGAT
780	GTAAGAAAAC	TGAATGTATT	TGATAAACAA	TTCATATACT	CACACACAAA	AAGAAATCAT
840	CTACTGCACA	TGAGATGAAG	TAGTCAATTA	TTTACATTCT	ACAAGCTACA	TTGATGAAAG
900	GGCACTTTCC	TACAAAATAT	CTGAATATAA	AAAAAATACA	TACCAAATAC	AAGTAGGGAG
960	AAGCCTACAA	TATTGGCATT	TTCTAGAATG	CACGGCGGGT	ATTTATCAAT	CCATGCCTAT
1020	AAAAAACCAA	AATTCCAACA	TCAACCCGTG	AAATATGACC	TATAATATAC	AACACACTCC
1080	GGAGCTAATC	TAGTTAAGAA	CAGTGCTCAA	CCTCAAACAA	CCAAACTATT	CCCAACCAAA
1140	AATACAAAGA	AATTGGGGCA	CAATAACATA	AAAGTAAAGC	ATTAAAAATA	CATTTTAGTA
1200	CTGTCATCCA	GGATCAGCTG	CATTAAATAA	TTGAATGATA	CAAAGTCAAG	TGGCTCTTAG
1260	TATGATGTGC	CACTCCCAAT	ATAATATTGA	AGTACAGGAG	TATTCAACGT	GCAAATACAC
1320	AATCATAAAT	TGAAGATGCA	TATTAATCAC	TGTGGTATGC	AAACAAACTA	AAAAACACCT
1380	GACACTATAA	AGGAAGGGAA	TGTCCAGGTT	TTATATGCTA	AATAGGTATG	TCACAGGATT
1440	ACAACATATC	AGTAGATATA .	AAGCTAATGG	TATCATGTTA	AGATGCTGGA	AGATACTTAA
1500	AGCTTGACAT	AACATTATCA	TCGAAGTATT	GAAATGAAAT	AAATGGAAAG	GTCAAGATAT

- 366 -

CAGAAATACA	AGTCAATATT	GAGATAGAAT	CTAGAAAGTC	CTACAAAAA	ATGCTAAAAG	1560
AGATGGGAGA	AGTGGCTCCA	GAATATAGGC	ATGATTCTCC	AGACTGTGGG	ATGATAATAC:	1620
TGTGTATAGC	TGCACTTGTG	ATAACCAAAT	TAGCAGCAGG	AGACAGATCA	GGTCTTACAG	1680
CAGTAATTAG	GAGGGCAAAC	AATGTCTTAA	AAAACGAAAT	AAAACGATAC	AAGGGCCTCA	1740
TACCAAAGGA	TATAGCTAAC	AGTTTTTATG	AAGTGTTTGA	AAAACACCCT	CATCTTATAG	1800
ATGTTTTCGT	GCACTTTGGC	ATTGCACAAT	CATCCACAAG	AGGGGGTAGT	AGAGTTGAAG	1860
GAATCTTTGC	AGGATTGTTT	ATGAATGCCT	ATGGTTCAGG	GCAAGTAATG	CTAAGATGGG	1920
GAGTTTTAGC	CAAATCTGTA	AAAAATATCA	TGCTAGGACA	TGCTAGTGTC	CAGGCAGAAA	1980
TGGAGCAAGT	TGTGGAAGTC	TATGAGTATG	CACAGAAGTT	GGGAGGAGAA	GCTGGATTCT	2040
ACCATATATT	GAACAATCCA	AAAGCATCAT	TGCTGTCATT	AACTCAATTT	CCCAACTTCT	2100
CAAGTGTGGT	CCTAGGCAAT	GCAGCAGGTC	TAGGCATAAT	GGGAGAGTAT	AGAGGTACAC	2160
CAAGAAACCA	GGATCTTTAT	GATGCAGCTA	AAGCATATGC	AGAGCAACTC	AAAGAAAATG	2220
GAGTAATAAA	CTACAGTGTA	TTAGACTTAA	CAGCAGAAGA	ATTGGAAGCC	ATAAAGCATC	2280
AACTCAACCC	CAAAGAAGAT	GATGTAGAGC	TTTAAGTTAA	CAAAAAATAC	GGGGCAAATA	2340
AGTCAACATG	GAGAAGTTTG	CACCTGAATT	TCATGGAGAA	GATGCAAATA	ACAAAGCTAC	2400
CAAATTCCTA	GAATCAATAA	AGGGCAAGTT	CGCATCATCC	AAAGATCCTA	AGAAGAAAGA	2460
TAGCATAATA	TCTGTTAACT	CAATAGATAT	AGAAGTAACT	AAAGAGAGCC	CGATAACATC	2520
TGGCACCAAC	ATCATCAATC	CAACAAGTGA	AGCCGACAGT	ACCCCAGAAA	CAAAAGCCAA	2580
CTACCCAAGA	AAACCCCTAG	TAAGCTTCAA	AGAAGATCTC	ACCCCAAGTG	ACAACCCTTT	2640
TTCTAAGTTG	TACAAGGAAA	CAATAGAAAC	ATTTGATAAC	aatgaagaag	AATCTAGCTA	2700
CTCATATGAA	GAGATAAATG	ATCARACARA	TGACAACATT	ACAGCAAGAC	TAGATAGAAT	2760
TGATGAAAAA	TTAAGTGAAA	TATTAGGAAT	GCTCCATACA	TTAGTAGTTG	CAAGTGCAGG	2820
ACCCACTTCA	GCTCGCGATG	GAATAAGAGA	TGCTATGGTT	GGTCTAAGAG	AAGAGATGA'I	2880
AGAAAAAATA	AGAGCGGAAG	CATTAATGAC	CAATGATAGG	TTAGAGGCTA	TGGCAAGACT	2940
TAGGAATGAG	GAAAGCGAAA	AAATGGCAAA	AGACACCTCA	GATGAAGTGT	CTCTTAATCC	3000
AACTTCCAAA	AAATTGAGTG	ACTTGTTGGA	AGACAACGAT	AGTGACAATG	ATCTATCACT	3060

- 367 -

TGATGATTTT	TGATCAGCGA	TCAACTCACT	CAGCAATCAA	CAACATCAAT	AAAACAGACA	3120
TCAATCCATT	GAATCAACTG	CCAGACCGAA	CAAACAAACG	TCCATCAGTA	GAACCACCAA	3180
CCAATCAATC	AACCAATTGA	TCAATCAGCA	ACCCGACAAA	ATTAACAATA	TAGTAACAAA	3240
AAAAGAACAA	GATGGGGCAA	ATATGGAAAC	ATACGTGAAC	AAGCTTCACG	AAGGCTCCAC	3300
ATACACAGCA	GCTGTTCAGT	ACAATGTTCT	AGAAAAAGAT	GATGATCCTG	CATCACTAAC	3360
AATATGGGTG	CCTATGTTCC	AGTCATCTGT	GCCAGCAGAC	TTGCTCATAA	AAGAACTTGC	3420
AAGCATCAAT	ATACTAGTGA	AGCAGATCTC	TACGCCCAAA	GGACCTTCAC	TACGAGTCAC	3480
GATTAACTCA	AGAAGTGCTG	TGCTGGCTCA	AATGCCTAGT	AATTTCATCA	TAAGCGCAAA	3540
TGTATCATTA	GATGAAAGAA	GCAAATTAGC	ATATGATGTA	ACTACACCTT	GTGAAATCAA	3600
AGCATGCAGT	CTAACATGCT	TAAAAGTAAA	AAGTATGTTA	ACTACAGTCA	AAGATCTTAC	3660
CATGAAGACA	TTCAACCCCA	CTCATGAGAT	CATTGCTCTA	TGTGAATTTG	AAAATATTAT	3720
GACATCAAAA	AGAGTAATAA	TACCAACCTA	TCTAAGATCA	ATTAGTGTCA	AGAACAAGGA	3780
TCTGAACTCA	CTAGAAAATA	TAGCAACCAC	CGAATTCAAA	AATGCTATCA	CCAATGCAAA	3840
AATTATTCCT	TATGCAGGAT	TAGTGTTAGT	TATCACAGTT	ACTGACAATA	AAGGAGCATT	3900
CAAATATATC	AAACCACAGA	GTCAATTTAT	AGTAGATCTT	GGTGCCTACC	TAGAAAAAGA	3960
GAGCATATAT	TATGTGACTA	CTAATTGGAA	GCATACAGCT	ACACGTTTTT	CAATCAAACC	4020
ACTAGAGGAT	TAAACTTAAT	TATCAACACT	GAATGACAGG	TCCACATATA	TCCTCAAACT	4080
ACACACTATA	TCCAAACATC	ATAAACATCT	ACACTACACA	CTTCATCACA	CAAACCAATC	4140
CCACTCAAAA	TCCAAAATCA	CTACCAGCCA	CTATCCGCTA	GACCTAGAGT	GCGAATAGGC	4200
AAATAAAACC	AAAATATGGG	GTAAATAGAC	ATTAGTTAGA	GTTCAATCAA	TCTTAACAAC	4260
CATTTATACC	GCCAATTCAA	CACATATACT	ATAAATCTTA	AAATGGGAAA	TACATCCATC	4320
ACAATAGAAC	TCACAAGCAA	ATTTTGGCCC	TATTTTACAC	TAATACATAT	GATCTTAACT	4380
CTAATCTTTT	TACTAATTAT	AATCACTATC	ATGATTGCAA	CACTAAATAA	GCTAAGTGAA	4440
CACAAAGCAT	TCTGCAACAA	AACTCTTGAA	CTAGGACAGA	TGTACCAAAT	CAACACACAG	4500
					CAAATCCAAT	4560
CCTCTCACAG	AGTCACGGTG	TCGCAAAACC	ACGCTAACCA	TCATGGTAGC	ATAGAGTAGT	4620

- 368 -

<b>TATTTAAAA</b>	TTAACATAAT	GATGAATTGT	TAGTATGAGA	TCAAAAACAA	CATTGGGGCA	4680
AATGCAACCA	TGTCCAAACA	CAAGAATCAA	CGCACTGCCA	GGACTCTAGA	AAAGACCTGG	4740
GATACTCTTA	ATCATCTAAT	TGTAATATCC	TCTTGTTTAT	ACAGATTAAA	TTTAAAATCT	4800
ATAGCACAAA	TAGCACTATC	AGTTTTGGCA	ATGATAATCT	CAACCTCTCT	CATAATTGCA	4860
GCCATAATAT	TCATCATCTC	TGCCAATCAC	AAAGTTACAC	TAACAACGGT	CACAGTTCAA	4920
ACAATAAAAA	ACCACACTGA	AAAAAACATC	ACCACCTACC	CTACTCAAGT	CTCACCAGAA	4980
AGGGTTAGTT	CATCCAAGCA	ACCCACAACC	ACATCACCAA	TCCACACAAG	TTCAGCTACA	5040
ACATCACCCA	ATACAAAATC	AGAAACACAC	CATACAACAG	CACAAACCAA	AGGCAGAACC	5100
ACCACTTCAA	CACAGACCAA	CAAGCCAAGC	ACAAAACCAC	GTCCAAAAAA	TCCACCAAAA	5160
AAAGATGATT	ACCATTTTGA	AGTGTTCAAC	TTCGTTCCCT	GCAGTATATG	TGGCAACAAT	5220
CAACTTTGCA	AATCCATCTG	CAAAACAATA	CCAAGCAACA	AACCAAAGAA	GAAACCAACC	5280
ATCAAACCCA	CAAACAAACC	AACCACCAAA	ACCACAAACA	AAAGAGACCC	AAAAACACCA	5340
GCCAAAACGA	CGAAAAAAGA	AACTACCACC	AACCCAACAA	AAAAACTAAC	CCTCAAGACC	5400
ACAGAAAGAG	ACACCAGCAC	CTCACAATCC	ACTGCACTCG	ACACAACCAC	ATTAAAACAC	5460
ACAGTCCAAC	AGCAATCCCT	CCTCTCAACC	ACCCCCGAAA	ACACACCCAA	CTCCACACAA	5520
ACACCCACAG	CATCCGAGCC	CTCCACACCA	AACTCCACCC	AAAAAACCCA	GCCACATGCT	5580
TAGTTATTCA	AAAACTACAT	CTTAGCAGAG	AACCGTGATC	TATCAAGCAA	GAACGAAATT'	5640
AAACCTGGGG	CAAATAACCA	TGGAGTTGAT	GATCCACAAG	TCAAGTGCAA	TCTTCCTAAC	5700
TCTTGCTATT	AATGCATTGT	ACCTCACCTC	AAGTCAGAAC	ATAACTGAGG	AGTTTTACCA	5760
ATCGACATGT	AGTGCAGTTA	GCAGAGGTTA	TTTTAGTGCT	TTAAGAACAG	GTTGGTATAC:	5820
TAGTGTCATA	ACAATAGAAT	TAAGTAATAT	AAAAGAAACC	AAATGCAATG	GAACTGACAC	5880
TAAAGTAAAA	CTTATGAAAC	AAGAATTAGA	TAAGTATAAG	AATGCAGTAA	CAGAATTACA	5940
GCTACTTATG	CAAAACACAC	CAGCTGTCAA	CAACCGGGCC	AGAAGAGAAG	CACCACAGTA	6000
TATGAACTAC	ACAATCAATA	CCACTAAAA	CCTAAATGTA	TCAATAAGCA	AGAAGAGGAA	6060
ACGAAGATTT	CTAGGCTTCT	TGTTAGGTGT	GGGATCTGCA	ATAGCAAGTG	GTATAGCTGT	6120
ATCAAAAGTT	CTACACCTTG	AAGGAGAAGT	GAACAAGATC	AAAAATGCTT	TGTTGTCTAC	6180

- 369 -

AAACAAAGC:	T GTAGTCAGT	TATCAAATG	GGTCAGTGT	r ttaaccagci	A AAGTGTTAGA	6240
TCTCAAGAA	TACATAAAT	ACCARTTAT:	r acccatagti	A AATCAACAG	GCTGTCGCAT	6300
CTCCAACAT:	F GAAACAGTT	TAGAATTCC	A GCAGAAGAA	C AGCAGATTGT	TGGAAATCAC	6360
CAGAGAATT	r agtgtcaate	CAGGTGTAA	AACACCTTTA	A AGCACTTACA	TGTTGACAAA	6420
CAGTGAGTT	A CTATCATTA	TCAATGATAT	GCCTATAACA	AATGATCAGA	TAATTAAAA	6480
GTCAAGCAA1	r GTTCAGATAG	TAAGGCAAC	AAGTTATTCC	ATCATGTCTA	TAATAAAGGA	6540
AGAAGTCCTT	GCATATGTTG	TACAGCTGCC	TATCTATGGT	GTAATAGATA	CACCTTGCTG	6600
GAAATTGCAC	ACATCGCCTC	TATGCACTAC	CAACATCAAA	GAAGGATCAA	ATATTTGTTT	6660
AACAAGGACT	GATAGAGGAT	GGTATTGTGA	TAATGCAGGA	TCAGTATCCT	TCTTTCCACA	6720
GGCTGACACT	TGTAAAGTAC	AGTCCAATCG	AGTATTTGT	GACACTATGA	ACAGTTTGAC	6780
ATTACCAAGT	GAAGTCAGCC	TTTGTAACAC	TGACATATTC	AATTCCAAGT	ATGACTGCAA	6840
AATTATGACA	TCAAAAACAG	ACATAAGCAG	CTCAGTAATT	ACTTCTCTTG	GAGCTATAGT	6900
GTCATGCTAT	GGTAAAACTA	AATGCACTGC	ATCCAACAAA	AATCGTGGGA	TTATAAAGAC	6960
ATTTTCTAAT	GGTTGTGACT	ATGTGTCAAA	CAAAGGAGTA	GATACTGTGT	CAGTGGGCAA	7020
CACTITATAC	TATGTAAACA	AGCTGGAAGG	CAAGAACCTT	TATGTAAAAG	GGGAACCTAT	7080
AATAAATTAC	TATGACCCTC	TAGTGTTTCC	TTCTGATGAG	TTTGATGCAT	CAATATCTCA	7140
AGTCAATGAA	AAAATCAATC	AAAGTTTAGC	TTTTATTCGT	AGATCTGATG	AATTACTACA	7200
TAATGTAAAT	ACTGGCAAAT	CTACTACAAA	TATTATGATA	ACTACAATTA	TTATAGTAAT	7260
CATTGTAGTA	TTGTTATCAT	TAATAGCTAT	TGGTTTACTG	TTGTATTGTA	AAGCCAAAAA	7320
CACACCAGTT	ACACTAAGCA	AAGACCAACT	AAGTGGAATC	AATAATATTG	CATTCAGCAA	7380
ATAGACAAAA	AACCACCTGA	TCATGTTTCA	ACAACAATCT	GCTGACCACC	AATCCCAAAT	7440
CAACTTACAA	CAAATATTTC	AACATCACAG	TACAGGCTGA	ATCATTTCCT	CACATCATGC	7500
TACCCACATA	ACTAAGCTAG	ATCCTTAACT	TATAGTTACA	TAAAAACCTC	AAGTATCACA	7560
ATCAACCACT	AAATCAACAC	ATCATTCACA	AAATTAACAG	CTGGGGCAAA	TATGTCGCGA	7620
AGAAATCCTT	GTAAATTTGA	GATTAGAGGT	CATTGCTTGA	ATGGTAGAAG	ATGTCACTAC	7680
AGTCATAATT	ACTTTGAATG	GCCTCCTCAT	GCATTACTAG	TGAGGCAAAA	CTTCATGTTA	7740

- 370 -

AACAAGATAC	TCAAGTCAAT	GGACAAAAGC	ATAGACACTT	TGTCTGAAAT	AAGTGGAGCT	7800
GCTGAACTGG	ATAGAACAGA	AGAATATGCT	CTTGGTATAG	TTGGAGTGCT	AGAGAGTTAC	7860
ATAGGATCTA	TAAACAACAT	AACAAAACAA	TCAGCATGTG	TTGCTATGAG	TAAACTTCTT	7920
ATTGAGATCA	ATAGTGATGA	CATTAAAAAG	CTTAGAGATA	ATGAAGAACC	CAATTCACCT	7980
AAGATAAGAG	TGTACAATAC	TGTTATATCA	TACATTGAGA	GCAATAGAAA	AAACAACAAG	8040
CAAACCATCC	ATCTGCTCAA	GAGACTACCA	GCAGACGTGC	TGAAGAAGAC	AATAAAGAAC	8100
ACATTAGATA	TCCACAAAAG	CATAACCATA	AGCAATCCAA	AAGAGTCAAC	TGTGAATGAT	8160
CAAAATGACC	AAACCAAAAA	TAATGATATT	ACCGGATAAA	TATCCTTGTA	GTATATCATC	8220
CATATTGATC	TCAAGTGAAA	GCATGGTTGC	TACATTCAAT	CATAAAAACA	TATTACAATT	8280
TAACCATAAC	TATTTGGATA	ACCACCAGCG	TTTATTAAAT	CATATATTTG	ATGAAATTCA	8340
TTGGACACCT	AAAAACTTAT	TAGATGCCAC	TCAACAATTT	CTCCAACATC	TTAACATCCC	8400
TGAAGATATA	TATACAGTAT	ATATATTAGT	GTCATAATGC	TTGACCATAA	CGACTCTATG	8460
TCATCCAACC	ATAAAACTAT	TTTGATAAGG	TTATGGGACA	AAATGGATCC	CATTATTAAT	8520
GGAAACTCTG	CTAATGTGTA	TCTAACTGAT	AGTTATTTAA	AAGGTGTTAT	CTCTTTTTCA	8580
GAGTGTAATG	CTTTAGGGAG	TTATCTTTTT	AACGGCCCTT	ATCTTAAAAA	TGATTACACC	8640
AACTTAATTA	GTAGACAAAG	CCCACTACTA	GAGCATATGA	ATCTTAAAAA	ACTAACTATA	8700
ACACAGTCAT	TAATATCTAG	ATATCATAAA	GGTGAACTGA	AATTAGAAGA	ACCAACTTAT	8760
TTCCAGTCAT	TACTTATGAC	ATATAAAAGT	ATGTCCTCGT	CTGAACAAAT	TGCTACAACT	8820
AACTTACTTA	AAAAAATAAT	ACGAAGAGCC	ATAGAAATAA	GTGATGTAAA	GGTGTACGCC	8880
ATCTTGAATA	AACTAGGATT	AAAGGAAAAG	GACAGAGTTA	AGCCCAACAA	TAATTCAGGT	8940
GATGAAAACT	CAGTACTTAC	AACCATAATT	AAAGATGATA	TACTTTCGGC	TGTGGAAAAC	9000
AATCAATCAT	ATACAAATTC	AGACAAAAGT	CACTCAGTAA	ATCAAAATAT	CACTATCAAA	9060
ACAACACTCT	TGAAAAATT	GATGTGTTCA	ATGCAACATC	CTCCATCATG	GTTAATACAC	9120
TGGTTCAATT	TATATACAAA	ATTAAATAAC	ATATTAACAC	AATATCGATC	AAATGAGGTA	9180
	GGTTTATATT					9240
AATCAATATG	GTTGTATCGT	TTATCATAAA	GGACTCAAAA	AAATCACAAC	TACTACTTAC	9300

- 371 -

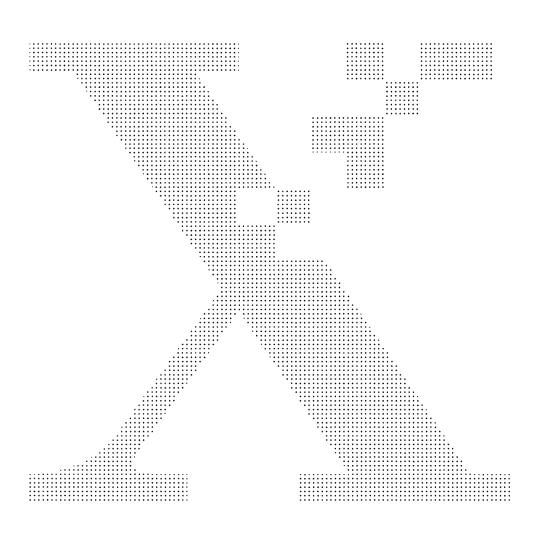
AATCAATTTI	TGACATGGAA	AGACATCAGO	: CTTAGCAGAT	TAAATGTTT	CTTAATTACT	9360
TGGATAAGTA	ATTGTTTAAA	TACATTAAAC	: AAAAGCTTAG	GGCTGAGAT	G TGGATTCAAT	9420
AATGTTGTGT	TATCACAATI	ATTTCTTTAT	GGAGATTGTA	TACTGAAATT	ATTTCATAAT	9480
GAAGGCTTCT	ACATAATAAA	AGAAGTAGAG	GGATTTATTA	. TGTCTTTAA1	TCTAAACATA	9540
ACAGAAGAAG	ATCAATTTAA	GAAACGATTI	TATAATAGCA	TGCTAAATAA	CATCACAGAT	9600
GCAGCTATTA	AGGCTCAAAA	GGACCTACTA	TCAAGAGTAT	GTCACACTTI	ATTAGACAAG	9660
ACAGTGTCTG	ATAATATCAT	AAATGGTAAA	TGGATAATCC	TATTAAGTAA	ATTTCTTAAA	9720
TTGATTAAGC	TTGCAGGTGA	TAATAATCTC	AATAACTTGA	GTGAGCTATA	TTTTCTCTTC	9780
AGAATCTTTG	GACATCCAAT	GGTCGATGAA	AGACAAGCAA	TGGATTCTGT	AAGAATTAAC	9840
TGTAATGAAA	CTAAGTTCTA	CTTATTAAGT	AGTCTAAGTA	CATTAAGAGG	TGCTTTCATT	9900
TATAGAATCA	TAAAAGGGTT	TGTAAATACC	TACAACAGAT	GGCCCACCTT	AAGGAATGCT	9960
ATTGTCCTAC	CTCTAAGATG	GTTAAACTAC	TATAAACTTA	ATACTTATCC	ATCTCTACTT	10020
GAAATCACAG	AAAATGATTT	GATTATTTTA	TCAGGATTGC	GGTTCTATCG	TGAGTTTCAT	10080
CTGCCTAAAA	AAGTGGATCT	TGAAATGATA	ATAAATGACA	AAGCCATTTC	ACCTCCAAAA	10140
GATCTAATAT	GGACTAGTTT	TCCTAGAAAT	TACATGCCAT	CACATATACA	AAATTATATA	10200
GAACATGAAA	AGTTGAAGTT	CTCTGAAAGC	GACAGATCGA	GAAGAGTACT	AGAGTATTAC	10260
TTGAGAGATA	ATAAATTCAA	TGAATGCGAT	CTATACAATT	GTGTAGTCAA	TCAAAGCTAT	10320
CTCAACAACT	CTAATCACGT	GGTATCACTA	ACTGGTAAAG	Aaagagagct	CAGTGTAGGT	10380
AGAATGTTTG	CTATGCAACC	AGGTATGTTT	AGGCAAATCC	AAATCTTAGC	AGAGAAAATG	10440
ATAGCTGAAA	ATATTTTACA	ATTCTTCCCT	GAGAGTTTGA	CAAGATATGG	TGATCTAGAG	10500
CTTCAAAAGA	TATTAGAATT	AAAAGCAGGA	ATAAGCAACA	AGTCAAATCG	TTATAATGAT	10560
AACTACAACA	ATTATATCAG	TAAATGTTCT	ATCATTACAG	ATCTTAGCAA	ATTCAATCAG	10620
GCATTTAGAT	ATGAAACATC	ATGTATCTGC	AGTGATGTAT	TAGATGAACT	GCATGGAGTA	10680
CAATCTCTGT	TCTCTTGGTT	GCATTTAACA	ATACCTCTTG	TCACAATAAT	ATGTACATAT	10740
AGACATGCAC	CTCCTTTCAT	AAAGGATCAT	GTTGTTAATC	TTAATGAGGT	TGATGAACAA	10800
AGTGGATTAT	ACAGATATCA	TATGGGTGGT	ATTGAGGGCT	GGTGTCAAAA	ACTGTGGACC	10860

- 372 -

ATTGAAGCTA	TATCATTATT	AGATCTAATA	TCTCTCAAAG	GGAAATTCTC	TATCACAGCT	10920
CTGATAAATG	GTGATAATCA	GTCAATTGAT	ATAAGCAAAC	CAGTTAGACT	TATAGAGGGT	10980
CAGACCCATG	CACAAGCAGA	TTATTTGTTA	GCATTAAATA	GCCTTAAATT	GTTATATAAA	11040
GAGTATGCAG	GTATAGGCCA	TAAGCTTAAG	GGAACAGAGA	CCTATATATC	CCGAGATATG	11100
CAGTTCATGA	GCAAAACAAT	CCAGCACAAT	GGAGTGTACT	ATCCAGCCAG	TATCARARA	11160
GTCCTGAGAG	TAGGTCCATG	GATAAACACG	ATACTTGATG	ATTTTAAAGT	TAGTTTAGAA	11220
TCTATAGGCA	GCTTAACACA	ggagttagaa	TACAGAGGAG	AAAGCTTATT	ATGCAGTTTA	11280
ATATTTAGGA	ACATTTGGTT	ATACAATCAA	ATTGCTTTGC	AACTCCGAAA	TCATGCATTA	11340
TGTAACAATA	AGCTATATTT	AGATATATTG	AAAGTATTAA	AACACTTAAA	AACTTTTTT	11400
AATCTTGATA	GCATTGATAT	GGCTTTATCA	TTGTATATGA	ATTTGCCTAT	GCTGTTTGGT	11460
GGTGGTGATC	CTAATTTGTT	ATATCGAAGC	TTTTATAGGA	GAACTCCAGA	CTTCCTTACA	11520
GAAGCTATAG	TACATTCAGT	GTTTGTGTTG	AGCTATTATA	CTGGTCACGA	TTTACAAGAT	11580
AAGCTCCAGG	ATCTTCCAGA	TGATAGACTG	AACAAATTCT	TGACATGTGT	CATCACATTT	11640
GATAAAAATC	CCAATGCCGA	GTTTGTAACA	TTGATGAGGG	ATCCACAGGC	TTTAGGGTC1	11700
GAAAGGCAAG	CTAAAATTAC	TAGTGAGATT	AATAGATTAG	CAGTAACAGA	AGTCTTAAGT	11760
ATAGCCCCAA	ACAAAATATT	TTCTAAAAGT	GCACAACATT	ATACTACCAC	TGAGATTGAT	11820
CTAAATGACA	TTATGCAAAA	TATAGAACCA	ACTTACCCTC	ATGGATTAAG	AGTTGTTTAT	11880
GAAAGTTTAC	CTTTTTATAA	AGCAGAAAA	ATAGTTAATC	TTATATCAGG	AACAAAATCC	11940
ATAACTAATA	TACTTGAAAA	AACATCAGCA	ATAGATACAA	CTGATATTAA	TAGGGCTACT	12000
GATATGATGA	GGAAAAATAT	AACTTTACTT	ATAAGGATAC	TTCCACTAGA	TTGTAACAAA	12060
GACAAAAGAG	AGTTATTAAG	TTTAGAAAAT	CTTAGTATAA	CTGAATTAAG	CAAGTATGTA	12120
AGAGAAAGAT	CTTGGTCATT	ATCCAATATA	GTAGGAGTAA	CATCGCCAAG	TATTATGTTC	12180
ACAATGAACA	TTAAATATAC	AACTAGCACT	ATAGCCAGTG	GTATAATAAT	AGAAAAATAT	12240
AATGTTAATA	GTTTAACTCG	TGGTGAAAGA	GGACCCACCA	AGCCATGGGT	AGGCTCATCC	12300
ACGCAGGAGA	AAAAAACAAT	GCCAGTGTAC	AACAGACAAG	TTTTAACCAA	AAAGCAAAGA	12360
GACCAAATAG	ATTTATTAGC	AAAATTAGAC	TGGGTATATG	CATCCATAGA	CAACAAAGAT	12420

# JP009927

B01 WO9813501.pdf 01/30/07 09:38 AM



- 373 -

GAATTCATGG	AAGAACTGAG	TACTGGAACA	CTTGGACTGT	CATATGAAAA	AGCCAAAAAG	12480
TTGTTTCCAC	AATATCTAAG	TGTCAATTAT	TTACACCGTT	TAACAGTCAG	TAGTAGACCA	12540
TGTGAATTCC	CTGCATCAAT	ACCAGCTTAT	AGAACAACAA	ATTATCATTT	TGATACTAGT	12600
CCTATCAATC	ATGTATTAAC	AGAAAAGTAT	GGAGATGAAG	ATATCGACAT	TGTGTTTCAA	12660
AATTGCATAA	GTTTTGGTCT	TAGCCTGATG	TCGGTTGTGG	AACAATTCAC	AAACATATGT	12720
CCTAATAGAA	TTATTCTCAT	ACCGAAGCTG	AATGAGATAC	ATTTGATGAA	ACCTCCTATA	12780
TTTACAGGAG	ATGTTGATAT	CATCAAGTTG	AAGCAAGTGA	TACAAAAGCA	GCACATGTTC	12840
CTACCAGATA	AAATAAGTTT	AACCCAATAT	GTAGAATTAT	TCTTAAGTAA	CAAAGCACTT	12900
AAATCTGGAT	CTCACATCAA	CTCTAATTTA	ATATTAGTAC	ATAAAATGTC	TGATTATTTT	12960
CATAATGCTT	ATATTTTAAG	TACTAATTTA	GCTGGACATT	GGATTCTGAT	TATTCAACTT	13020
ATGAAAGATT	CAAAAGGTAT	TTTTGAAAAA	GATTGGGGAG	AGGGGTACAT	AACTGATCAT	13080
ATGTTCATTA	ATTTGAATGT	TTTCTTTAAT	GCTTATAAGA	CTTATTTGCT	ATGTTTTCAT	13140
AAAGGTTATG	GTAAAGCAAA	ATTAGAATGT	GATATGAACA	CTTCAGATCT	TCTTTGTGTT	13200
TTGGAGTTAA	TAGACAGTAG	CTACTGGAAA	TCTATGTCTA	AAGTTTTCCT	AGAACAAAA	13260
GTCATAAAAT	ACATAGTCAA	TCAAGACACA	AGTTTGCGTA	GAATAAAAGG	CTGTCACAGT	13320
TTTAAGTTGT	GGTTTTTAAA	ACGCCTTAAT	AATGCTAAAT	TTACCGTATG	CCCTTGGGTT	13380
GTTAACATAG	ATTATCACCC	AACACACATG	AAAGCTATAT	TATCTTACAT	AGATTTAGTT	13440
AGAATGGGGT	TAATAAATGT	AGATAAATTA	ACCATTAAAA	ATAAAAACAA	ATTCAATGAT	13500
GAATTTTACA	CATCAAATCT	CTTTTACATT	AGTTATAACT	TTTCAGACAA	CACTCATTTG	13560
CTAACAAAAC	AAATAAGAAT	TGCTAATTCA	GAATTAGAAG	AATATTAATA	CAAACTATAT	13620
CACCCAACCC	CAGAAACTTT	AGAAAATATG	TCATTAATTC	CTGTTAAAAG	TAATAATAGT	13680
AACAAACCTA .	AATTTTGTAT	AAGTGGAAAT	ACCGAATCTA	TGATGATGTC	AACATTCTCT	13740
AGTAAAATGC .	ATATTAAATC	TTCCACTGTT	ACCACAAGAT	TCAATTATAG	CAAACAAGAC	13800
					AGGTAATACA	13860
GCAAAATCTA 2						13920
GCATCACTTT I	ATTGCATGCT	TCCTTGGCAT	CATGTCAATA	GATTTAACTT	TGTATTTAGT	13980

- 374 -

TCCAC	AGGAT	GCAAGATCAG	TATAGAGTAT	ATTTTAAAAG	ATCTTAAGAT	TAAGGACCCC	14040
AGTTG	TATAG	CATTCATAGG	TGAAGGAGCT	GGTAACTTAT	TATTACGTAC	GGTAGTAGAA	14100
CTTCA	TCCAG	ACATAAGATA	CATTTACAGA	<b>AGTTTAAAA</b> G	ATTGCAATGA	TCATAGTTTA	14160
CCTAT	TG <b>A</b> AT	TTCTAAGGTT	ATACAACGGG	CATATAAACA	TAGATTATGG	TGAGAATTTA	14220
ACCAT	TCCTG	CTACAGATGC	AACTAATAAC	ATTCATTGGT	CTTATTTACA	TATAAAATTT	14280
GCAGA	ACCTA	TTAGCATCTT	TGTCTGCGAT	GCTGAATTAC	CTGTTACAGC	CAATTGGAGT	14340
TAAAA	TATAA	TTGAATGGAG	TAAGCATGTA	AGAAAGTGCA	AGTACTGTTC	TTCTGTAAAT	14400
AGATG	CATTT	TAATTGCAAA	ATATCATGCT	CAAGATGACA	TTGATTTCAA	ATTAGATAAC:	14460
ATTAC	TATAT	TAAAAACTTA	CGTGTGCCTA	GGTAGCAAGT	TAAAAGGATC	TGAAGTTTAC	14520
TTAAT	CCTTA	CAATAGGCCC	TGCAAATATA	CTTCCTGTTT	TTGATGTTGT	ACAAAATGCT	14580
TTAAA	GATAC	TTTCAAGAAC	TAAAAATTTC	ATTATGCCTA	AAAAAACTGA	CAAGGAATCT	14640
ATCGA	TGCAA	ATATTAAAAG	CTTAATACCT	TTCCTTTGTT	ACCCTATAAC	AAAAAAGGA	14700
ATTAA	GACTT	CATTGTCAAA	ATTGAAGAGT	GTAGTTAATG	GAGATATATT	ATCATATTCT	14760
ATAGC	TGGAC	GTAATGAAGT	ATTCAGCAAC	AAGCTTATAA	ACCACAAGCA	TATGAATATC	14820
CTAAA	ATGGC	TAGATCATGT	TTTAAATTTT	AGATCAGCTG	AACTTAATTA	CAATCATTTA	14880
TACAT	GATAG	AGTCCACATA	TCCTTACTTA	AGTGAATTGT	TAAATAGTTT	AACAACCAAT	14940
GAGCT	CAAGA	AGCTGATTAA	AATAACAGGT	AGTGTGCTAT	ACAACCTTCC	CAACGAACAG	15000
TAGTT	AAAA	TATCATTAAC	AAGTTTGGTC	AAATTTAGAT	GCTAACACAT	CATTATATTA	15060
TAGTT.	ATTAA	AGAATATACA	AACTTTTCAA	TAATTTAGCA	TATTGATTCC	AAAATTATCA	15120
TTTTA	GTCTT	AAGGGGTTAA	ATAAAAGTCT	AAAACTAACA	ATTATACATG	TGCATTCACA	15180
ACACA	ACGAG	ACATTAGTTT	TTGACACTTT	TTTTCTCGT			15219

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

- 375 -

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

- Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15
- Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly
  20 25 30
- Ser Tyr Leu Phe Asn Gly Pro Tyr Leu Lys Asn Asp Tyr Thr Asn Leu 35 40 45
- Ile Ser Arg Gln Ser Pro Leu Leu Glu His Met Asn Leu Lys Lys Leu 50 55 60
- Thr Ile Thr Gln Ser Leu Ile Ser Arg Tyr His Lys Gly Glu Leu Lys 65 70 75 80
- Leu Glu Glu Pro Thr Tyr Phe Gln Ser Leu Leu Met Thr Tyr Lys Ser 85 90 95
- Met Ser Ser Glu Gln Ile Ala Thr Thr Asn Leu Leu Lys Lys Ile 100 105 110
- Ile Arg Arg Ala Ile Glu Ile Ser Asp Val Lys Val Tyr Ala Ile Leu 115 120 125
- Asn Lys Leu Gly Leu Lys Glu Lys Asp Arg Val Lys Pro Asn Asn Asn 130 135 140
- Ser Gly Asp Glu Asn Ser Val Leu Thr Thr Ile Ile Lys Asp Asp Ile
  145 150 155 160
- Leu Ser Ala Val Glu Asn Asn Gln Ser Tyr Thr Asn Ser Asp Lys Ser 165 170 175
- His Ser Val Asn Gln Asn Ile Thr Ile Lys Thr Thr Leu Leu Lys Lys 180 185 190
- Leu Met Cys Ser Met Gln His Pro Pro Ser Trp Leu Ile His Trp Phe 195 200 205
- Asn Leu Tyr Thr Lys Leu Asn Asn Ile Leu Thr Gln Tyr Arg Ser Asn 210 215 220
- Glu Val Lys Ser His Gly Phe Ile Leu Ile Asp Asn Gln Thr Leu Ser 225 230 235 240
- Gly Phe Gln Phe Ile Leu Asn Gln Tyr Gly Cys Ile Val Tyr His Lys

- 376 -

				245					250					255	
Gly	Leu	Lys	Lys 260	Ile	Thr	Thr	Thr	Thr 265	Tyr	Asn	Gln	Phe	Leu 270	Thr	Trp
Lys	Asp	11e 275	Ser	Leu	Ser	Arg	Leu 280	Asn	Val	Сув	Leu	Ile 285	Thr	Trp	Ile
Ser	Asn 290	Cys	Leu	Asn	Thr	Leu 295	Asn	Lys	Ser	Leu	Gly 300	Leu	Arg	Сув	Gly
Phe 305	Asn	Asn	Val	Val	Leu 310	Ser	Gln	Leu	Phe	Leu 315	Tyr	Gly	Asp	Сув	11e 320
Leu	Lys	Leu	Phe	His 325	Asn	Glu	Gly	Phe	Tyr 330	Ile	Ile	Lys	Glu	Val 335	Glu
Gly	Phe	Ile	Met 340	Ser	Leu	Ile	Leu	Asn 345	Ile	Thr	Glu	Glu	Asp 350	Gln	Phe
Lys	Lys	Arg 355	Phe	Tyr	Asn	Ser	Met 360	Leu	Asn	Asn	Ile	Thr 365	Asp	Ala	Ala
Ile	Lys 370	Ala	Gln	Lys	Ąsp	Leu 375	Leu	Ser	Arg	Val	Сув 380	His	Thr	Leu	Leu
Asp 385	Lys	Thr	Val	Ser	<b>Asp</b> 390	Asn	Ile	Ile	Asn	Gly 395	Lys	Trp	Ile	Ile	Leu 400
Leu	Ser	Lys	Phe	Leu 405	Lys	Leu	Ile	Lys	Leu 410	Ala	Gly	Asp	Asn	Asn 415	Leu
Asn	Asn	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
Met	Val	Asp 435	Glu	Arg	Gln	Ala	Met 440	Asp	Ser	Val	Arg	Ile 445	Asn	Сув	Asn
Glu	Thr 450	Lys	Phe	Tyr	Leu	Leu 455	Ser	Ser	Leu	Ser	Thr 460	Leu	Arg	Gly	Ala
Phe 465	Ile	Tyr	Arg	Ile	Ile 470	Lys	Gly	Phe	Val	Asn 475	Thr	Tyr	Asn	Arg	Trp 480
Pro	Thr	Leu	Arg	Asn 485	Ala	Ile	Val	Leu	Pro 490	Leu	Arg	Trp	Leu	Asn 495	Tyr
Tyr	Lys	Leu	Asn 500	Thr	Tyr	Pro	Ser	Leu 505	Leu	Glu	Ile	Thr	Glu 510	Asn	Asp
Leu	Ile	Ile 515	Leu	Ser	Gly	Leu	Arg 520	Phe	Tyr	Arg	Glu	Phe 525	His	Leu	Pro

- 377 -

Lys	Lys 530	Va:	l Ası	Lev	ı Glu	535		Ile	a Ası	n Asp	540		a Il	e Se	r Pro
Pro 545	Lys	As <sub>l</sub>	) Let	ı Ile	550		s Ser	Phe	Pro	555		Ту	r Me	t Pro	Ser 560
His	Ile	Glr	n Asr	565	Ile	Glu	ı His	Glu	1 Lys 570		Lys	Phe	e Sei	: Gl: 575	s Ser
Asp	Arg	Ser	580	Arg	Val	Leu	. Glu	Tyr 585		Leu	Arg	, As	590		Phe
		595	i				600					605	<b>i</b>		a Asn
Asn	Ser 610	Asn	His	Val	Val	Ser 615		Thr	Gly	Lys	Glu 620		, Glu	Leu	Ser
625					630					635					Gln 640
				645			Ala		650					655	
			660				Asp	665					670		
		675					Lys 680					685			
	690					695	Ser				700				
705					710		Thr			715					720
Asp				725					730					735	
			740				Сув	745					750		
Ile		755					760					765			
	/70					775					780				
Trp :	Chr	Ile	Glu	Ala	Ile 790	Ser	Leu	Leu		Leu 795	Ile	Ser	Leu	Lys	<b>Gly</b> 800

- 378 -

Lys	Phe	Ser	Ile	Thr 805	Ala	Leu	Ile	Asn	Gly 810	Ąsp	Asn	Gln	Ser	11e 815	<b>qa</b> A
Ile	Ser	Lys	Pro 820	Val	Arg	Leu	Ile	Glu 825	Gly	Gln	Thr	His	Ala 830	Gln	Ala
qaA	Tyr	Leu 835	Leu	Ala	Leu	Asn	Ser 840	Leu	Lys	Leu	Leu	Tyr 845	Lys	Glu	Tyr
Ala	Gly 850	Ile	Gly	His	Lys	Leu 855	Lys	Gly	Thr	Glu	Thr 860	Tyr	Ile	Ser	Arg
Asp 865	Met	Gln	Phe	Met	Ser 870	Lys	Thr	Ile	Gln	His 875	Asn	Gly	Val	Tyr	Tyr 880
Pro	Ala	Ser	Ile	Lys 885	ГЛЯ	Val	Leu	Arg	Val 890	Gly	Pro	Trp	Ile	Asn 895	Thr
Ile	Leu	Asp	900	Phe	Lys	Val	Ser	Leu 905	Glu	Ser	Ile	Gly	Ser 910	Leu	Thr
Gln	Glu	Leu 915	Glu	Tyr	Arg	Gly	Glu 920	Ser	Leu	Leu	Сув	Ser 925	Leu	Ile	Phe
Arg	Asn 930	Ile	Trp	Leu	Tyr	Asn 935	Gln	Ile	Ala	Leu	Gln 940	Leu	Arg	Asn	His
Ala 945	Leu	Сув	Asn	Asn	Lys 950	Leu	Tyr	Leu	Asp	11e 955	Leu	Lys	Val	Leu	<b>L</b> ув 960
His	Leu	Lys	Thr	Phe 965	Phe	Asn	Leu	Авр	Ser 970	Ile	Asp	Met	Ala	<b>Leu</b> 975	Ser
Leu	Tyr	Met	Asn 980	Leu	Pro	Met	Leu	Phe 985	Gly	Gly	Gly	Asp	Pro 990	Asn	Leu
Leu	Tyr	Arg 995	Ser	Phe	Tyr	Arg	Arg 1000		Pro	Asp	Phe	Leu 100!		Glu	Ala
Ile	Val 101		Ser	Val	Phe	Val 101		Ser	Tyr	Tyr	Thr 102	_	His	Asp	Leu
Gln 102	<b>Asp</b> 5	Lys	Leu	Gln	Asp 103		Pro	Авр	Asp	Arg 103		Asn	Lys	Phe	Leu 104
Thr	Сув	Val	Ile	Thr 104		Asp	Lys	Asn	Pro 105		Ala	Glu	Phe	Val 105	
Leu	Met	Arg	Asp 106		Gln	Ala	Leu	Gly 106		Glu	Arg	Gln	Ala 107	_	Ile
Thr	Ser	Glu	Ile	Asn	Arg	Leu	Ala	Val	Thr	Glu	Val	Leu	Ser	Ile	Ala

- 379 -

		107	5				1080	)				108	5		
Pro	Asn 1090	_	Ile	Phe	Ser	Lys 1095		Ala	Gln	His	Tyr 110		Thr	Thr	Glu
Ile 1105		Leu	Asn	Asp	Ile 1110		Gln	Asn	Ile	Glu 111:		Thr	Tyr	Pro	His 1120
Gly	Leu	Arg	Val	Val 1125		Glu	Ser	Leu	Pro 1130		Tyr	Lys	Ala	Glu 113	_
Ile	Val	Asn	Leu 1140	Ile	Ser	Gly	Thr	Lys 1145		Ile	Thr	Asn	Ile 1150		Glu
Lys	Thr	Ser 1155		Ile	Asp	Thr	Thr 1160	_	Ile	Asn	Arg	Ala 1165		Авр	Met
Met	Arg 1170		Asn	Ile	Thr	<b>Leu</b> 1175		Ile	Arg	Ile	Leu 118(		Leu	Asp	Сув
Asn 1185		yab	Lys	Arg	Glu 1190		Leu	Ser	Leu	Glu 1195		Leu	Ser	Ile	Thr 1200
Glu	Leu	Ser	Lys	Tyr 1205		Arg	Glu	Arg	Ser 1210		Ser	Leu	Ser	Asn 1215	
Val	Gly	Val	Thr 1220	Ser	Pro	Ser	Ile	Met 1225		Thr	Met	Asn	lle 1230	_	Tyr
Thr	Thr	Ser 1235		Ile	Ala	Ser	Gly 1240		Ile	Ile	Glu	Lys 1245	-	Asn	Val
Asn	Ser 1250		Thr	Arg	Gly	Glu 1255	_	Gly	Pro	Thr	Lys 1260		Trp	Val	Gly
Ser 1265		Thr	Gln	Glu	Lув 1270	_	Thr	Met	Pro	Val 1275	_	Asn	Arg	Gln	Val 1280
Leu	Thr	Lys	Lys	Gln 1285		Asp	Gln	Ile	Asp 1290		Leu	Ala	Lys	Leu 1295	_
Trp	Val	Tyr	Ala 1300	Ser	Ile	Asp	Asn	Lys 1305		Glu	Phe	Met	Glu 1310		Leu
Ser		Gly 1315		Leu	Gly		Ser 1320		Glu	Lys	Ala	Lys 1325		Leu	Phe
Pro	Gln 1330	_	Leu	Ser		Asn 1335	-	Leu	His	_	Leu 1340		Val	Ser	Ser
Arg 1345		Сув	Glu	Phe	Pro 1350		Ser	Ile		Ala 1355		Arg	Thr		Asn 1360

- Tyr His Phe Asp Thr Ser Pro Ile Asn His Val Leu Thr Glu Lys Tyr 1365 1370 1375
- Gly Asp Glu Asp Ile Asp Ile Val Phe Gln Asn Cys Ile Ser Phe Gly 1380 1385 1390
- Leu Ser Leu Met Ser Val Val Glu Gln Phe Thr Asn Ile Cys Pro Asn 1395 1400 1405
- Arg Ile Ile Leu Ile Pro Lys Leu Asn Glu Ile His Leu Met Lys Pro 1410 1415 1420
- Pro Ile Phe Thr Gly Asp Val Asp Ile Ile Lys Leu Lys Gln Val Ile 1425 1430 1435 1440
- Gln Lys Gln His Met Phe Leu Pro Asp Lys Ile Ser Leu Thr Gln Tyr 1445 1450 1455
- Val Glu Leu Phe Leu Ser Asn Lys Ala Leu Lys Ser Gly Ser His Ile 1460 1465 1470
- Asn Ser Asn Leu Ile Leu Val His Lys Met Ser Asp Tyr Phe His Asn 1475 1480 1485
- Ala Tyr Ile Leu Ser Thr Asn Leu Ala Gly His Trp Ile Leu Ile Ile 1490 1495 1500
- Gln Leu Met Lys Asp Ser Lys Gly Ile Phe Glu Lys Asp Trp Gly Glu 1505 1510 1515 1520
- Gly Tyr Ile Thr Asp His Met Phe Ile Asn Leu Asn Val Phe Phe Asn 1525 1530 1535
- Ala Tyr Lys Thr Tyr Leu Leu Cys Phe His Lys Gly Tyr Gly Lys Ala 1540 1545 1550
- Lys Leu Glu Cys Asp Met Asn Thr Ser Asp Leu Leu Cys Val Leu Glu 1555 1560 1565
- Leu Ile Asp Ser Ser Tyr Trp Lys Ser Met Ser Lys Val Phe Leu Glu 1570 1580
- Gln Lys Val Ile Lys Tyr Ile Val Asn Gln Asp Thr Ser Leu Arg Arg 1585 1590 1595 1600
- Ile Lys Gly Cys His Ser Phe Lys Leu Trp Phe Leu Lys Arg Leu Asn 1605 1610 1615
- Asn Ala Lys Phe Thr Val Cys Pro Trp Val Val Asn Ile Asp Tyr His 1620 1625 1630

- 381 -

- Pro Thr His Met Lys Ala Ile Leu Ser Tyr Ile Asp Leu Val Arg Met 1635 1640 1645
- Gly Leu Ile Asn Val Asp Lys Leu Thr Ile Lys Asn Lys Asn Lys Phe 1650 1655 1660
- Asn Asp Glu Phe Tyr Thr Ser Asn Leu Phe Tyr Ile Ser Tyr Asn Phe 1665 1670 1675 1680
- Ser Asp Asn Thr His Leu Leu Thr Lys Gln Ile Arg Ile Ala Asn Ser 1685 1690 1695
- Glu Leu Glu Asp Asn Tyr Asn Lys Leu Tyr His Pro Thr Pro Glu Thr 1700 1705 1710
- Leu Glu Asn Met Ser Leu Ile Pro Val Lys Ser Asn Asn Ser Asn Lys 1715 1720 1725
- Pro Lys Phe Cys Ile Ser Gly Asn Thr Glu Ser Met Met Ser Thr 1730 1735 1740
- Phe Ser Ser Lys Met His Ile Lys Ser Ser Thr Val Thr Thr Arg Phe 1745 1750 1755 1760
- Asn Tyr Ser Lys Gln Asp Leu Tyr Asn Leu Phe Pro Ile Val Val Ile 1765 1770 1775
- Asp Lys Ile Ile Asp His Ser Gly Asn Thr Ala Lys Ser Asn Gln Leu 1780 1785 1790
- Tyr Thr Thr Ser His Gln Thr Ser Leu Val Arg Asn Ser Ala Ser 1795 1800 1805
- Leu Tyr Cys Met Leu Pro Trp His His Val Asn Arg Phe Asn Phe Val 1810 1815 1820
- Phe Ser Ser Thr Gly Cys Lys Ile Ser Ile Glu Tyr Ile Leu Lys Asp 1825 1830 1835 1840
- Leu Lys Ile Lys Asp Pro Ser Cys Ile Ala Phe Ile Gly Glu Gly Ala 1845 1850 1855
- Gly Asn Leu Leu Leu Arg Thr Val Val Glu Leu His Pro Asp Ile Arg 1860 1865 1870
- Tyr Ile Tyr Arg Ser Leu Lys Asp Cys Asn Asp His Ser Leu Pro Ile 1875 1880 1885
- Glu Phe Leu Arg Leu Tyr Asn Gly His Ile Asn Ile Asp Tyr Gly Glu 1890 1895 1900
- Asn Leu Thr Ile Pro Ala Thr Asp Ala Thr Asn Asn Ile His Trp Ser

- 382 -

1905				1910	)				191	5				1920
Tyr Leu	His	Ile	Lys 192		Ala	Glu	Pro	11e 193		Ile	Phe	Val	Cys 1935	_
Ala Glu	Leu	Pro 1940		Thr	Ala	Asn	Trp 194		Lys	Ile	Ile	Ile 195		Trp
Ser Lys	His 195		Arg	Lys	Сув	Lув 1960		Сув	Ser	Ser	Val 196		Arg	Cys
Ile Leu 197		Ala	Lys	Tyr	His 197		Gln	Asp	Asp	Ile 198	_	Phe	Lys	Leu
Asp Asn 1985	Ile	Thr	Ile	Leu 1990		Thr	Tyr	Val	Сув 199!		Gly	Ser	Lys	Leu 2000
Lys Gly	Ser	Glu	Val 200		Leu	Ile	Leu	Thr 2010		Gly	Pro	Ala	Asn 2015	
Leu Pro	Val	Phe 202		Val	Val	Gln	Asn 202		Lys	Leu	Ile	Leu 2031		Arg
Thr Lys	Asn 203		Ile	Met	Pro	Lys 2040	_	Thr	Asp	Lys	Glu 204!		Ile	Asp
Ala Asn 205		ГУв	Ser	Leu	Ile 205		Phe	Leu	Сув	Tyr 2060		Ile	Thr	Lys
Lys Gly 2065	Ile	Lys	Thr	Ser 207		Ser	Lys	Leu	Lys 207		Val	Val	Asn	Gly 2080
Asp Ile	Leu	Ser	Tyr 208		Ile	Ala	Gly	Arg 2090		Glu	Val	Phe	Ser 2095	
Lys Leu	Ile	Asn 210		Lys	His	Met	Asn 210		Leu	Lys	Trp	Leu 2110	_	His
Val Leu	Asn 211		Arg	Ser	Ala	Glu 2120		Asn	Tyr	Asn	His 212		Tyr	Met
Ile Glu 213		Thr	Tyr	Pro	Tyr 213		Ser	Glu	Leu	Leu 2140		Ser	Leu	Thr
Thr Asn 2145	Glu	Leu	Lys	Lys 2150		Ile	Lys	Ile	Thr 215		Ser	Val	Leu	Tyr 2160
Asn Leu	Pro	Asn	Glu 216										•	

(2) INFORMATION FOR SEQ ID NO:33:

- 383 -

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: RNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACGGGAAAAA	AATGCGTACT	ACAAACTTGC	ACATTCGAAA	AAAATGGGGC	AAATAAGAAC	60
TTGATAAGTG	CTATTTAAGT	CTAACCTTTT	CAATCAGAAA	TGGGGTGCAA	TTCACTGAGC	120
ATGATAAAGG	TTAGATTACA	AAATTTATTT	GACAATGACG	AAGTAGCATT	GTTAAAAATA	180
ACATGTTATA	CTGATAAATT	AATTCTTCTG	ACCAATGCAT	TAGCCAAAGC	AGCAATACAT	240
ACAATTAAAT	TAAACGGCAT	AGTTTTATA	CATGTTATAA	CAAGCAGTGA	AGTGTGCCCT	300
GATAACAATA	TTGTAGTGAA	ATCTAACTTT	ACAACAATGC	CAATACTACA	AAATGGAGGA	360
TACATATGGG	<b>AATTGATT</b> GA	GTTGACACAC	TGCTCTCAAT	TAAACGGTTT	AATGGATGAT	420
aattgtgaaa	TCAAATTTTC	TAAAAGACTA	AGTGACTCAG	TAATGACTAA	TTATATGAAT	480
CAAATATCTG	ACTTACTTGG	GCTTGATCTC	AATTCATGAA	TTATGTTTAG	TCTAATTCAA	540
TAGACATGTG	TTTATTACCA	TTTTAGTTAA	TATAAAAACT	CATCAAAGGG	AAATGGGGCA	600
AATAAACTCA	CCTAATCAAT	CAAACCATGA	GCACTACAAA	TGACAACACT	ACTATGCAAA	660
GATTGATGAT	CACAGACATG	AGACCCCTGT	CAATGGATTC	AATAATAACA	TCTCTTACCA	720
AAGAAATCAT	CACACACAAA	TTCATATACT	TGATAAACAA	TGAATGTATT	GTAAGAAAAC	780
TTGATGAAAG	ACAAGCTACA	TTTACATTCT	TAGTCAATTA	TGAGATGAAG	CTACTGCACA	840
AAGTAGGGAG	TACCAAATAC	AAAAAATACA	CTGAATATAA	TACAAAATAT	GGCACTTTCC	900
CCATGCCTAT	ATTTATCAAT	CACGGCGGGT	TTCTAGAATG	TATTGGCATT	AAGCCTACAA	960
AACACACTCC	TATAATATAC	AAATATGACC	TCAACCCGTG	AATTCCAACA	AAAAAACCAA	1020
CCCAACCAAA	CCAAACTATT	CCTCAAACAA	CAGTGCTCAA	TAGTTAAGAA	GGAGCTAATC	1080
CATTTTAGTA	ATTAAAATTA	AAAGTAAAGC	CAATAACATA	AATTGGGGCA	AATACAAAGA	1140
TGGCTCTTAG	CAAAGTCAAG	TTGAATGATA	CATTARATAA	GGATCAGCTG	CTGTCATCCA	1200

- 384 -

GCAAATACAC	TATTCAACGT	AGTACAGGAG	ATAATATTGA	CACTCCCAAT	TATGATGTGC	1260
AAAAACACCT	AAACAAACTA	TGTGGTATGC	TATTAATCAC	TGAAGATGCA	AATCATAAAT	1320
TCACAGGATT	AATAGGTATG	TTATATGCTA	TGTCCAGGTT	AGGAAGGGAA	GACACTATAA	1380
AGATACTTAA	AGATGCTGGA	TATCATGTTA	AAGCTAATGG	AGTAGATATA	ACAACATATC	1440
GTCAAGATAT	AAATGGAAAG	GAAATGAAAT	TCGAAGTATT	AACATTATCA	AGCTTGACAT	1500
CAGAAATACA	AGTCAATATT	GAGATAGAAT	CTAGAAAGTC	CTACAAAAA	ATGCTAAAAG	1560
AGATGGGAGA	AGTGGCTCCA	GAATATAGGC	ATGATTCTCC	AGACTGTGGG	ATGATAATAC	1620
TGTGTATAGC	TGCACTTGTG	ATAACCAAAT	TAGCAGCAGG	AGACAGATCA	GGTCTTACAG	1680
CAGTAATTAG	GAGGGCAAAC	AATGTCTTAA	AAAACGAAAT	AAAACGATAC	AAGGGCCTCA	1740
TACCAAAGGA	TATAGCTAAC	AGTTTTTATG	AAGTGTTTGA	AAAACACCCT	CATCTTATAG	1800
ATGTTTTCGT	GCACTTTGGC	ATTGCACAAT	CATCCACAAG	AGGGGGTAGT	agagttgaag	1860
GAATCTTTGC	AGGATTGTTT	ATGAATGCCT	ATGGTTCAGG	GCAAGTAATG	CTAAGATGGG	1920
GAGTTTTAGC	CAAATCTGTA	AAAAATATCA	TGCTAGGACA	TGCTAGTGTC	CAGGCAGAAA	1980
TGGAGCAAGT	TGTGGAAGTC	TATGAGTATG	CACAGAAGTT	GGGAGGAGAA	GCTGGATTCT	2040
ACCATATATT	GAACAATCCA	AAAGCATCAT	TGCTGTCATT	AACTCAATTT	CCCAACTTCT	2100
CAAGTGTGGT	CCTAGGCAAT	GCAGCAGGTC	TAGGCATAAT	GGGAGAGTAT	AGAGGTACAC	2160
CAAGAAACCA	GGATCTTTAT	GATGCAGCTA	AAGCATATGC	AGAGCAACTC	AAAGAAAATG	2220
GAGTAATAAA	CTACAGTGTA	TTAGACTTAA	CAGCAGAAGA	ATTGGAAGCC	ATAAAGCATC	2280
AACTCAACCC	CAAAGAAGAT	GATGTAGAGC	TTTAAGTTAA	CAAAAAATAC	GGGGCAAATA	2340
AGTCAACATG	GAGAAGTTTG	CACCTGAATT	TCATGGAGAA	GATGCAAATA	ACAAAGCTAC	2400
CAAATTCCTA	GAATCAATAA	AGGGCAAGTT	CGCATCATCC	AAAGATCCTA	AGAAGAAAGA	2460
TAGCATAATA	TCTGTTAACT	CAATAGATAT	AGAAGTAACT	AAAGAGAGCC	CGATAACATC	2520
TGGCACCAAC	ATCATCAATC	CAACAAGTGA	AGCCGACAGT	ACCCCAGAAA	CAAAAGCCAA	2580
CTACCCAAGA	AAACCCCTAG	TAAGCTTCAA	AGAAGATCTC	ACCCCAAGTG	ACAACCCTTT	2640
TTCTAAGTTG	TACAAGGAAA	CAATAGAAAC	ATTTGATAAC	AATGAAGAAG	AATCTAGCTA	2700
CTCATATGAA	GAGATAAATG	ATCAAACAAA	TGACAACATT	ACAGCAAGAC	TAGATAGAAT	2760

- 385 -

TGATGAAAA	TTAAGTGAAA	TATTAGGAAT	GCTCCATAC	TTAGTAGTT	CAAGTGCAGG	2820
ACCCACTTCA	GCTCGCGATG	GAATAAGAGA	TGCTATGGTT	GGTCTAAGAG	AAGAGATGAT	2880
AGAAAAAATA	AGAGCGGAAG	CATTAATGAC	CANTGATAGG	TTAGAGGCT	TGGCAAGACT	2940
TAGGAATGAG	GAAAGCGAAA	AAATGGCAAA	AGACACCTCA	GATGAAGTGI	CTCTTAATCC	3000
AACTTCCAAA	AAATTGAGTG	ACTTGTTGGA	AGACAACGAT	AGTGACAATG	ATCTATCACT	3060
TGATGATTTT	TGATCAGCGA	TCAACTCACT	CAGCAATCAA	CAACATCAAT	AAAACAGACA	3120
TCAATCCATT	GAATCAACTG	CCAGACCGAA	CAAACAAACG	TCCATCAGTA	GAACCACCAA	3180
CCAATCAATC	AACCAATTGA	TCAATCAGCA	ACCCGACAAA	ATTAACAATA	TAGTAACAAA	3240
AAAAGAACAA	GATGGGGCAA	ATATGGAAAC	ATACGTGAAC	AAGCTTCACG	AAGGCTCCAC	3300
ATACACAGCA	GCTGTTCAGT	ACAATGTTCT	agaaaaagat	GATGATCCTG	CATCACTAAC	3360
AATATGGGTG	CCTATGTTCC	AGTCATCTGT	GCCAGCAGAC	TTGCTCATAA	AAGAACTTGC	3420
AAGCATCAAT	ATACTAGTGA	AGCAGATCTC	TACGCCCAAA	GGACCTTCAC	TACGAGTCAC	3480
					TAAGCGCAAA	3540
					GTGAAATCAA	3600
					AAGATCTTAC	3660
					TATTATAAAA	3720
					AGAACAAGGA	3780
					CCAATGCAAA	3840
					AAGGAGCATT	3900
					TAGAAAAAGA	3960
	TATGTGACTA					4020
					TCCTCAAACT	4080
					CAAACCAATC	4140
					GCGAATAGGT TCTTAACAAC	4200
CATTTATACC						4260
	COURTIONS	CACAIATACT	AAAATCTTA	AAATGGGAAA	TACATCCATC	4320

- 386 -

ACAATAGAAT	TCACAAGCAA	ATTTTGGCCC	TATTTTACAC	TAATACATAT	GATCTTAACT	4380
CTAATCTTTT	TACTAATTAT	AATCACTATT	ATGATTGCAA	TACTAAATAA	GCTAAGTGAA	4440
CATAAAGCAT	TCTGTAACAA	AACTCTTGAA	CTAGGACAGA	TGTATCAAAT	CAACACATAG	4500
AGTTCTACCA	TTATGCTGTG	TCAAATTATA	ATCCTGTATA	TATAAACAAA	CARATCCART	4560
CTTCTCACAG	AGTCATGGTG	TCGCAAAACC	ACGCTAACTA	TCATGGTAGC	ATAGAGTAGT	4620
TATTTAAAA	TTAACATAAT	GATGAATTGT	TAGTATGAGA	TCAAAAACAA	CATTGGGGCA	4680
AATGCAACCA	TGTCCAAACA	CAAGAATCAA	CGCACTGCCA	GGACTCTAGA	AAAGACCTGG	4740
GATACTCTTA	ATCATCTAAT	TGTAATATCC	TCTTGTTTAT	ACAGATTAAA	TTTAAAATCT	4800
ATAGCACAAA	TAGCACTATC	AGTTTTGGCA	ATGATAATCT	CAACCTCTCT	CATAATTGCA	4860
GCCATAATAT	TCATCATCTC	TGCCAATCAC	AAAGTTACAC	TAACAACGGT	CACAGTTCAA	4920
ACAATAAAAA	ACCACACTGA	AAAAAACATC	ACCACCTACC	CTACTCAAGT	CTCACCAGAA	4980
AGGGTTAGTT	CATCCAAGCA	ACCCACAACC	ACATCACCAA	TCCACACAAG	TTCAGCTACA	5040
ACATCACCCA	ATACAAAATC	AGAAACACAC	CATACAACAG	CACAAACCAA	AGGCAGAACC	5100
ACCACTTCAA	CACAGACCAA	CAAGCCAAGC	ACAAAACCAC	GTCCAAAAAA	TCCACCAAAA	5160
			TTCGTTCCCT			5220
			CCAAGCAACA			5280
			ACCACAAACA			5340
			AACCCAACAA			5400
			ACTGCACTCG			5460
			ACCCCCGAAA			5520
			AACTCCACCC			5580
					GAACGAAATT	5640
					TCTTCCTAAC	5700
					AGTTTTACCA	5760
					GTTGGTATAC	5820
IAGIGICATA	ACAATAGAAT	TAAGTAATAT	AAAAGAAACC	AAATGCAATG	GAACTGACAC	5880

- 387 -

TAAAGTAAAA	CTTATGAAAC	AAGAATTAGA	TAAGTATAAG	AATGCAGTAA	CAGAATTACA	5940
GCTACTTATG	CAAAACACAC	CAGCTGTCAA	CAACCGGGCC	: AGAAGAGAAG	CACCACAGTA	6000
TATGAACTAC	ACARTCARTA	CCACTAAAAA	CCTAAATGTA	TCANTANGCA	AGAAGAGGAA	6060
ACGAAGATTT	CTAGGCTTCT	TGTTAGGTGT	GGGATCTGCA	ATAGCAAGTG	GTATAGCTGT	6120
ATCAAAAGTT	CTACACCTTG	AAGGAGAAGT	GAACAAGATO	AAAAATGCTT	TGTTGTCTAC	6180
AAACAAAGCT	GTAGTCAGTT	TATCAAATGG	GGTCAGTGTT	TTAACCAGCA	AAGTGTTAGA	6240
TCTCAAGAAT	TACATAAATA	ACCAATTATT	ACCCATAGTA	AATCAACAGA	GCTGTCGCAT	6300
CTCCAACATT	GAAACAGTTA	TAGAAȚTCCA	GCAGAAGAAC	AGCAGATTGT	TGGAAATCAC	6360
CAGAGAATTT	AGTGTCAATG	CAGGTGTAAC	AACACCTTTA	AGCACTTACA	TGTTGACAAA	6420
CAGTGAGTTA	CTATCATTAA	TCAATGATAT	GCCTATAACA	AATGATCAGA	TAATTAAAA	6480
GTCAAGCAAT	GTTCAGATAG	TAAGGCAACA	AAGTTATTCC	ATCATGTCTA	TAATAAAGGA	6540
AGAAGTCCTT	GCATATGTTG	TACAGCTGCC	TATCTATGGT	GTAATAGATA	CACCTTGCTG	6600
GAAATTGCAC	ACATCGCCTC	TATGCACTAC	CAACATCAAA	GAAGGATCAA	ATATTTGTTT	6660
AACAAGGACT	GATAGAGGAT	GGTATTGTGA	TAATGCAGGA	TCAGTATCCT	TCTTTCCACA	6720
GGCTGACACT	TGTAAAGTAC	AGTCCAATCG	AGTATTTTGT	GACACTATGA	ACAGTTTGAC	6780
ATTACCAAGT	GAAGTCAGCC	TTTGTAACAC	TGACATATTC	AATTCCAAGT	ATGACTGCAA	6840
AATTATGACA	TCAAAAACAG	ACATAAGCAG	CTCAGTAATT	ACTTCTCTTG	GAGCTATAGT	6900
GTCATGCTAT	GGTAAAACTA	AATGCACTGC	ATCCAACAAA	AATCGTGGGA	TTATAAAGAC	6960
ATTTTCTAAT	GGTTGTGACT	ATGTGTCAAA	CAAAGGAGTA	GATACTGTGT	CAGTGGGCAA	7020
CACTITATAC	TATGTAAACA	AGCTGGAAGG	CAAGAACCTT	TATGTAAAAG	GGGAACCTAT	7080
AATAAATTAC						7140
					AATTACTACA	7200
					TTATAGTAAT	7260
					AAGCCAAAAA	7320
					CATTCAGCAA	7380
ATAGACAAAA .	AACCACCTGA	TCATGTTTCA	ACAACAATCT	GCTGACCACC	AATCCCAAAT	7440

- 388 -

CAACTTACAA	CAAATATTTC	AACATCACAG	TACAGGCTGA	ATCATTTCCT	CACATCATGC	7500
TACCCACATA	ACTAAGCTAG	ATCCTTAACT	TATAGTTACA	TAAAAACCTC	AAGTATCACA	7560
ATCAACCACT	AAATCAACAC	ATCATTCACA	AAATTAACAG	CTGGGGCAAA	TATGTCGCGA	7620
AGAAATCCTT	GTAAATTTGA	GATTAGAGGT	CATTGCTTGA	ATGGTAGAAG	ATGTCACTAC	7680
AGTCATAATT	ACTTTGAATG	GCCTCCTCAT	GCATTACTAG	TGAGGCAAAA	CTTCATGTTA	7740
AACAAGATAC	TCAAGTCAAT	GGACAAAAGC	ATAGACACTT	TGTCTGAAAT	AAGTGGAGCT	7800
GCTGAACTGG	ATAGAACAGA	AGAATATGCT	CTTGGTATAG	TTGGAGTGCT	AGAGAGTTAC	7860
ATAGGATCTA	TAAACAACAT	AACAAAACAA	TCAGCATGTG	TTGCTATGAG	TAAACTTCTT	7920
ATTGAGATCA	ATAGTGATGA	CATTAAAAAG	CTTAGAGATA	ATGAAGAACC	CAATTCACCT	7980
AAGATAAGAG	TGTACAATAC	TGTTATATCA	TACATTGAGA	GCAATAGAAA	AAACAACAAG	8040
CAAACCATCC	ATCTGCTCAA	GAGACTACCA	GCAGACGTGC	TGAAGAAGAC	AATAAAGAAC	8100
ACATTAGATA	TCCACAAAAG	CATAACCATA	AGCAATCCAA	AAGAGTCAAC	TGTGAATGAT	8160
CAAAATGACC	AAACCAAAAA	TAATGATATT	ACCGGATAAA	TATCCTTGTA	GTATATCATC	8220
CATATTGATC	TCAAGTGAAA	GCATGGTTGC	TACATTCAAT	CATAAAAACA	TATTACAATT	8280
TAACCATAAC	TATTTGGATA	ACCACCAGCG	TTTATTAAAT	CATATATTTG	ATGAAATTCA	8340
TTGGACACCT	AAAAACTTAT	TAGATGCCAC	TCAACAATTT	CTCCAACATC	TTAACATCCC	8400
TGAAGATATA	TATACAGTAT	ATATATTAGT	GTCATAATGC	TTGACCATAA	CGACTCTATG	8460
TCATCCAACC	ATAAAACTAT	TTTGATAAGG	TTATGGGACA	AAATGGATCC	CATTATTAAT	8520
GGAAACTCTG	CTAATGTGTA	TCTAACTGAT	AGTTATTTAA	AAGGTGTTAT	CTCTTTTCA	8580
GAGTGTAATG	CTTTAGGGAG	TTATCTTTTT	AACGGCCCTT	ATCTTAAAAA	TGATTACACC	8640
AACTTAATTA	GTAGACAAAG	CCCACTACTA	GAGCATATGA	ATCTTAAAAA	ACTARCTATA	8700
ACACAGTCAT	TAATATCTAG	ATATCATAAA	GGTGAACTGA	aattagaaga	ACCAACTTAT	8760
TTCCAGTCAT	TACTTATGAC	ATATAAAAGT	ATGTCCTCGT	CTGAACAAAT	TGCTACAACT	8820
AACTTACTTA	TAATAAAAA	ACGAAGAGCC	ATAGAAATAA	GTGATGTAAA	GGTGTACGCC	8880
ATCTTGAATA	AACTAGGATT	AAAGGAAAAG	GACAGAGTTA	AGCCCAACAA	TAATTCAGGT	8940
GATGAAAACT	CAGTACTTAC	AACTATAATT	AAAGATGATA	TACTTTCGGC	TGTGGAAAAC	9000

- 389 -

AATCAATCAT	ATACAAATTC	AGACAAAAGT	CACTCAGTAA	ATCAAAATAT	CACTATCAAA	9060
ACAACACTCT	TGAAAAAATT	GATGTGTTCA	ATGCAACATC	CTCCATCAT	GTTAATACAC	9120
TGGTTCAATT	TATATACAAA	ATTAAATAAC	ATATTAACAC	AATATCGATC	: AAATGAGGTA	9180
AAAAGTCATG	GGTTTATATT	AATAGATAAT	CAAACTTTAA	GTGGTTTTCA	GTTTATTTTA	9240
AATCAATATG	GTTGTATCGT	TTATCATAAA	GGACTCAAAA	AAATCACAAC	TACTACTTAC	9300
AATCAATTTT	TGACATGGAA	AGACATCAGC	CTTAGCAGAT	TAAATGTTTG	CTTAATTACT	9360
TGGATAAGTA	ATTGTTTAAA	TACATTAAAC	AAAAGCTTAG	GGCTGAGATG	TGGATTCAAT	9420
AATGTTGTGT	TATCACAATT	ATTTCTTTAT	GGAGATTGTA	TACTGAAATT	ATTTCATAAT	9480
GAAGGCTTCT	ACATAATAAA	agaagtagag	GGATTTATTA	TGTCTTTAAT	TCTAAACATA	9540
ACAGAAGAAG	ATCAATTTAG	GAAACGATTT	TATAATAGCA	TGCTAAATAA	CATCACAGAT	9600
GCAGCTATTA	AGGCTCAAAA	GGACCTACTA	TCAAGAGTAT	GTCACACTTT	ATTAGACAAG	9660
ACAGTGTCTG	ATAATATCAT	Aaatggtaaa	TGGATAATCC	TATTAAGTAA	ATTTCTTAAA	9720
TTGATTAAGC	TTGCAGGTGA	TAATAATCTC	AATAACTTGA	GTGAGCTATA	TTTTCTCTTC	9780
AGAATCTTTG	GACATCCAAT	GGTCGATGAA	AGACAAGCAA	TGGATTCTGT	AAGAATTAAC	9840
TGTAATGAAA	CTAAGTTCTA	CTTATTAAGT	AGTCTAAGTA	CATTAAGAGG	TGCTTTCATT	9900
TATAGAATCA	TAAAAGGGTT	TGTAAATACC	TACAACAGAT	GGCCCACCTT	AAGGAATGCT	9960
ATTGTCCTAC	CTCTAAGATG	GTTAAACTAC	TATAAACTTA	ATACTTATCC	ATCTCTACTT	10020
GAAATCACAG	AAAATGATTT	GATTATTITA	TCAGGATTGC	GGTTCTATCG	TGAGTTTCAT	10080
CTGCCTAAAA	AAGTGGATCT	TGAAATGATA	ATAAATGACA	AAGCCATTTC	ACCTCCAAAA	10140
GATCTAATAT	GGACTAGTTT	TCCTAGAAAT	TACATGCCAT	CACATATACA	AAATTATATA	10200
GAACATGAAA	AGTTGAAGTT	CTCTGAAAGC	GACAGATCGA	GAAGAGTACT	AGAGTATTAC	10260
TTGAGAGATA .	ATAAATTCAA	TGAATGCGAT	CTATACAATT	GTGTAGTCAA	TCAAAGCTAT	10320
CTCAACAACT	CTAATCACGT	GGTATCACTA	actggtaaag	aaagagagct	CAGTGTAGGT	10380
AGAATGTTTG	CTATGCAACC	AGGTATGTTT	AGGCAAATCC	AAATCTTAGC	AGAGAAAATG	10440
ATAGCTGAAA	ATATTTTACA	ATTCTTCCCT	GAGAGTTTGA	CAAGATATGG	TGATCTAGAG	10500
CTTCAAAAGA :	FATTAGAATT .	AAAAGCAGGA	ATAAGCAACA	AGTCAAATCG	TTATAATGAT	10560

- 390 -

AACTACAACA	ATTATATCAG	TAAATGTTCT	ATCATTACAG	ATCTTAGCAA	ATTCAATCAG	10620
GCATTTAGAT	ATGAAACATC	ATGTATCTGC	AGTGATGTAT	TAGATGAACT	GCATGGAGTA	10680
CAATCTCTGT	TCTCTTGGTT	GCATTTAACA	ATACCTCTTG	TCACAATAAT	ATGTACATAT	10740
AGACATGCAC	CTCCTTTCAT	AAAGGATCAT	GTTGTTAATC	TTAATGAGGT	TGATGAACAA	10800
AGTGGATTAT	ACAGATATCA	TATGGGTGGT	ATTGAGGGCT	GGTGTCAAAA	ACTGTGGACC	10860
ATTGAAGCTA	TATCATTATT	AGATCTAATA	TCTCTCAAAG	GGAAATTCTC	TATCACAGCT	10920
CTGATAAATG	GTGATAATCA	GTCAATTGAT	ATAAGCAAAC	CAGTTAGACT	TATAGAGGGT	10980
CAGACCCATG	CACAAGCAGA	TTATTTGTTA	GCATTAAATA	GCCTTAAATT	GTTATATAAA	11040
GAGTATGCAG	GTATAGGCCA	TAAGCTTAAG	GGAACAGAGA	CCTATATATC	CCGAGATATG	11100
CAGTTCATGA	GCAAAACAAT	CCAGCACAAT	GGAGTGTACT	ATCCAGCCAG	TATCARARA	11160
GTCCTGAGAG	TAGGTCCATG	GATAAACACG	ATACTTGATG	ATTTTAAAGT	TAGTTTAGAA	11220
			TACAGAGGAG			11280
			ATTGCTTTGC			11340
			AAAGTATTAA			11400
			TTGTATATGA			11460
			TTTTATAGGA			11520
			AGCTATTATA			11580
			AACAAATTCT			11640
			TTGATGAGGG			11700
			AATAGATTAG			11760
			GCACAACATT			11820
CTAAATGACA						11880
GAAAGTTTAC						11940
ATAACTAATA						12000
GATATGATGA						12060
GACAAAAGAG	AGITATTAAG	TITAGAAAAT	CTTAGTATAA	CTGAATTAAG	CAAGTATGTA	12120

- 391 -

AGAGAAAGAT	CTTGGTCATT	ATCCAATATA	GTAGGAGTAA	CATCGCCAAG	TATTATGTTC	12180
ACAATGGACA	TTAAATATAC	AACTAGCACT	ATAGCCAGTG	GTATAATAAT	AGAAAAATAT	12240
AATGTTAATA	GTTTAACTCG	TGGTGAAAGA	GGACCCACCA	AGCCATGGGT	AGGCTCATCC	12300
ACGCAGGAGA	AAAAAACAAT	GCCAGTGTAC	AACAGACAAG	TTTTAACCAA	AAAGCAAAGA	12360
GACCAAATAG	ATTTATTAGC	AAAATTAGAC	TGGGTATATG	CATCCATAGA	CAACAAAGAT	12420
GAATTCATGG	AAGAACTGAG	TACTGGAACA	CTTGGACTGT	CATATGAAAA	AGCCAAAAAG	12480
TTGTTTCCAC	AATATCTAAG	TGTCAATTAT	TTACACCGTT	TAACAGTCAG	TAGTAGACCA	12540
TGTGAATTCC	CTGCATCAAT	ACCAGCTTAT	AGAACAACAA	ATTATCATTT	TGATACTAGT	12600
CCTATCAATC	ATGTATTAAC	AGAAAAGTAT	GGAGATGAAG	ATATCGACAT	TGTGTTTCAA	12660
AATTGCATAA	GTTTTGGTCT	TAGCCTGATG	TCGGTTGTGG	AACAATTCAC	AAACATATGT	12720
CCTAATAGAA	TTATTCTCAT	ACCGAAGCTG	AATGAGATAC	ATTTGATGAA	ACCTCCTATA	12780
TTTACAGGAG	ATGTTGATAT	CATCAAGTTG	AAGCAAGTGA	TACAAAAGCA	GCACATGTTC	12840
CTACCAGATA	AAATAAGTTT	AACCCAATAT	GTAGAATTAT	TCTTAAGTAA	CAAAGCACTT	12900
AAATCTGGAT	CTCACATCAA	CTCTAATTTA	ATATTAGTAC	ATAAAATGTC	TGATTATTTT	12960
CATAATGCTT	ATATTTTAAG	TACTAATTTA	GCTGGACATT	GGATTCTGAT	TATTCAACTT	13020
ATGAAAGATT	CAAAAGGTAT	TTTTGAAAAA	GATTGGGGAG	AGGGGTACAT	AACTGATCAT	13080
ATGTTCATTA	ATTTGAATGT	TTTCTTTAAT	GCTTATAAGA	CTTATTTGCT	ATGTTTTCAT	13140
AAAGGTTATG	GTAAAGCAAA	ATTAGAATGT	GATATGAACA	CTTCAGATCT	TCTTTGTGTT	13200
TTGGAGTTAA	TAGACAGTAG	CTACTGGAAA	TCTATGTCTA	AAGTTTTCCT	AGAACAAAAA	13260
GTCATAAAAT	ACATAGTCAA	TCAAGACACA	AGTTTGCGTA	GAATAAAAGG	CTGTCACAGT	13320
TTTAAGTTGT	GGTTTTTAAA	ACGCCTTGAT	AATGCTAAAT	TTACCGTATG	CCCTTGGGTT	13380
GTTAACATAG	ATTATCACCC	AACACACATG	AAAGCTATAT	TATCTTACAT	AGATTTAGTT	13440
agaatggggt	TAATAAATGT	AGATAAATTA	ACCATTAAAA	ATAAAAACAA	ATTCAATGAT	13500
GAATTTTACA	CATCAAATCT	CTTTTACATT	AGTTATAACT	TTTCAGACAA	CACTCATTTG	13560
CTAACAAAAC	AAATAAGAAT	TGCTAATTCA	GAATTAGAAG	ATATTATA	CAAACTATAT	13620
CACCCAACCC	CAGAAACTTT	AGAAAATATG	TCATTAATTC	CTGTTAAAAG	Taataatagt	13680

- 392 -

AACAAACCTA	AATTTTGTAT	AAGTGGAAAT	ACCGAATCTA	TGATGATGTC	AACATTCTCT	13740
AGTAAAATGC	ATATTAAATC	TTCCACTGTT	ACCACAAGAT	TCAATTATAG	CAAACAAGAC	13800
TTGTACAATT	TATTTCCAAT	TGTTGTGATA	GACAAGATTA	TAGATCATTC	AGGTAATACA	13860
GCAAAATCTA	ACCAACTTTA	CACCACCACT	TCACATCAGA	CATCTTTAGT	AAGGAATAGT	13920
GCATCACTTT	ATTGCATGCT	TCCTTGGCAT	CATGTCAATA	GATTTAACTT	TGTATTTAGT	13980
TCCACAGGAT	GCAAGATCAG	TATAGAGTAT	attttaaaag	ATCTTAAGAT	TAAGGACCCC	14040
AGTTGTATAG	CATTCATAGG	TGAAGGAGCT	GGTAACTTAT	TATTACGTAC	GGTAGTAGAA	14100
CTTCATCCAG	ACATAAGATA	CATTTACAGA	AGTTTAAAAG	ATTGCAATGA	TCATAGTTTA	14160
CCTATTGAAT	TTCTAAGGTT	ATACAACGGG	CATATAAACA	TAGATTATGG	TGAGAATTTA	14220
ACCATTCCTG	CTACAGATGC	AACTAATAAC	ATTCATTGGT	CTTATTTACA	TATAAAATTT	14280
GCAGAACCTA	TTAGCATCTT	TGTCTGCGAT	GCTGAATTAC	CTGTTACAGC	CAATTGGAGT	14340
AAAATTATAA	TTGAATGQAG	TAAGCATGTA	AGAAAGTGCA	AGTACTGTTC	TTCTGTAAAT	14400
AGATGCATTT	TAATTGCAAA	ATATCATGCT	CAAGATGACA	TTGATTTCAA	ATTAGATAAC	14460
ATTACTATAT	TAAAAACTTA	CGTGTGCCTA	GGTAGCAAGT	TAAAAGGATC	TGAAGTTTAC	14520
TTAATCCTTA	CAATAGGCCC	TGCAAATATA	CTTCCTGTTT	TTGATGTTGT	ACAAAATGCT	14580
AAATTGATAC	TTTCAAGAAC	TAAAAATTTC	ATTATGCCTA	AAAAAACTGA	CAAGGAATCI	14640
ATCGATGCAG	TTATTAAAAG	CTTAATACCT	TTCCTTTGTT	ACCCTATAAC	AAAAAAAGGA	14700
ATTAAGACTT	CATTGTCAAA	ATTGAAGAGT	GTAGTTAATG	GAGATATATT	ATCATATTCT	14760
ATAGCTGGAC	GTAATGAAGT	ATTCAGCAAC	AAGCTTATAA	ACCACAAGCA	TATGAATATC	14820
CTAAAATGGC	TAGATCATGT	TTTAAATTTT	AGATCAGCTG	AACTTAATTA	CAATCATTTA	14880
TACATGATAG	AGTCCACATA	TCCTTACTTA	AGTGAATTGT	TAAATAGTTT	AACAACCAAT	14940
GAGCTCAAGA	AGCTGATTAA	AATAACAGGT	AGTGTGCTAT	ACAACCTTCC	CAACGAACAG	15000
TAGTTTAAAA	TATCATTAAC	AAGTTTGGTC	AAATTTAGAT	GCTAACACAT	CATTATATTA	15060
TAGTTATTAA	AAAATATACA	AACTTTTCAA	TAATTTAGCA	TATTGATTCC	AAAATTATCA	15120
TTTTAGTCTT	AAGGGGTTAA	ATAAAAGTCT	AAAACTAACA	ATTATACATG	TGCATTCACA	15180
ACACAACGAG	ACATTAGTTT	TTGACACTTT	TTTTCTCGT			15219

- 393 -

- (2) INFORMATION FOR SEQ ID NO:34:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2166 amino acids
    - (B) TYPE: amino acid
    - (C) STRANDEDNESS:
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:
  - Met Asp Pro Ile Ile Asn Gly Asn Ser Ala Asn Val Tyr Leu Thr Asp 1 5 10 15
  - Ser Tyr Leu Lys Gly Val Ile Ser Phe Ser Glu Cys Asn Ala Leu Gly 20 25 30
  - Ser Tyr Leu Phe Asn Gly Pro Tyr Leu Lys Asn Asp Tyr Thr Asn Leu 35 40 45
  - Ile Ser Arg Gln Ser Pro Leu Leu Glu His Met Asn Leu Lys Lys Leu 50 55 60
  - Thr Ile Thr Gln Ser Leu Ile Ser Arg Tyr His Lys Gly Glu Leu Lys 65 70 75 80
  - Leu Glu Glu Pro Thr Tyr Phe Gln Ser Leu Leu Met Thr Tyr Lys Ser 85 90 95
  - Met Ser Ser Ser Glu Gln Ile Ala Thr Thr Asn Leu Leu Lys Lys Ile 100 105 110
  - Ile Arg Arg Ala Ile Glu Ile Ser Asp Val Lys Val Tyr Ala Ile Leu 115 120 125
  - Asn Lys Leu Gly Leu Lys Glu Lys Asp Arg Val Lys Pro Asn Asn Asn 130 135 140
  - Ser Gly Asp Glu Asn Ser Val Leu Thr Thr Ile Ile Lys Asp Asp Ile 145 150 155 160
  - Leu Ser Ala Val Glu Asn Asn Gln Ser Tyr Thr Asn Ser Asp Lys Ser 165 170 175
  - His Ser Val Asn Gln Asn Ile Thr Ile Lys Thr Thr Leu Leu Lys Lys 180 185 190

- 394 -

Leu	Met	Сув 195	Ser	Met	Gln	His	Pro 200	Pro	Ser	Trp	Leu	11e 205	His	Trp	Phe
Asn	Leu 210	Tyr	Thr	Lys	Leu	Asn 215	Asn	Ile	Leu	Thr	Gln 220	Tyr	Arg	Ser	Asn
Glu 225	Val	Lys	Ser	His	Gly 230	Phe	Ile	Leu	Ile	Авр 235	Asn	Gln	Thr	Leu	Ser 240
Gly	Phe	Gln	Phe	11e 245	Leu	Asn	Gln	Tyr	Gly 250	Сув	Ile	Val	Tyr	His 255	Lys
Gly	Leu	Lys	Lys 260	Ile	Thr	Thr	Thr	Thr 265	Tyr	Asn	Gln	Phe	Leu 270	Thr	Trp
Lys	Asp	11e 275	Ser	Leu	Ser	Arg	Leu 280	Asn	Val	Сув	Leu	Ile 285	Thr	Trp	Ile
Ser	Asn 290	Сув	Leu	Asn	Thr	Leu 295	Asn	Lys	Ser	Leu	Gly 300	Leu	Arg	Сув	Gly
Phe 305	Asn	Asn	Val	Val	Leu 310	Ser	Gln	Leu	Phe	Leu 315	Tyr	Gly	Asp	Сув	11e 320
Leu	Lys	Leu	Phe	His 325	Asn	Glu	Gly	Phe	Tyr 330	Ile	Ile	Lys	Glu	Val 335	Glu
Gly	Phe	Ile	Met 340	Ser	Leu	Ile	Leu	Asn 345	Ile	Thr	Glu	Glu	<b>Asp</b> 350	Gln	Phe
Arg	Lys	Arg 355	Phe	Tyr	Asn	Ser	Met 360	Leu	Asn	Asn	Ile	Thr 365	Asp	Ala	Ala
Ile	Lys 370	Ala	Gln	Lys	qaA	Leu 375	Leu	Ser	Arg	Val	Сув 380	His	Thr	Leu	Leu
Asp 385	Lys	Thr	Val	Ser	<b>qaA</b>	Asn	Ile	Ile	Asn	Gly 395	Lys	Trp	Ile	Ile	<b>Leu</b>
Leu	Ser	Lys	Phe	Leu 405	Lys	Leu	Ile	Lys	Leu 410	Ala	Gly	Asp	Asn	Asn 415	Leu
Asn	Asn	Leu	Ser 420	Glu	Leu	Tyr	Phe	Leu 425	Phe	Arg	Ile	Phe	Gly 430	His	Pro
Met	Val	Авр 435	Glu	Arg	Gln	Ala	Met 440	Авр	Ser	Val	Arg	Ile 445	Asn ,	Сув	Asn
	450				Leu	455					460			_	
Phe	Ile	Tyr	Arg	Ile	Ile	Lys	Gly	Phe	Val	Asn	Thr	Tyr	Asn	Arg	Trp

- 395 -

465	•				470	ı				47	5				480
Pro	Thi	: Lei	ı Arg	485		Ile	Val	Let	1 Pro		u Arg	j Trj	) Let	1 Ası 495	Tyr
Tyr	Lys	Lev	Asn 500		Tyr	Pro	Ser	505		ı Glı	ı Ile	Thi	510		a Asp
Leu	Ile	11e 515		Ser	Gly	Leu	Arg 520		Туг	Arg	g Glu	Phe 525		Leu	Pro
Lys	Lys 530	Val	Asp	Leu	Glu	Met 535		Ile	· Asr	As <sub>I</sub>	540		Ile	Ser	Pro
Pro 545	Lys	Asp	Leu	Ile	Trp 550	Thr	Ser	Phe	Pro	555		Tyr	Met	Pro	Ser 560
His	Ile	Gln	Asn	Tyr 565	Ile	Glu	His	Glu	Lys 570		Lys	Phe	Ser	Glu 575	
Asp	Arg	Ser	Arg 580	Arg	Val	Leu	Glu	Tyr 585	Tyr	Leu	Arg	Asp	Asn 590		Phe
Asn	Glu	Суя 595	yab	Leu	Tyr	Asn	Сув 600	Val	Val	Asn	Gln	Ser 605	Tyr	Leu	Asn
Asn	Ser 610	Asn	His	Val	Val	Ser 615	Leu	Thr	Gly	Lys	Glu 620	Arg	Glu	Leu	Ser
Val 625	Gly	Arg	Met	Phe	Ala 630	Met	Gln	Pro	Gly	Met 635	Phe	Arg	Gln	Ile	Gln 640
Ile	Leu	Ala	Glu	Lys 645	Met	Ile	Ala	Glu	Asn 650	Ile	Leu	Gln	Phe	Phe 655	Pro
Glu	Ser	Leu	Thr 660	Arg	Tyr	Gly	Asp	Leu 665	Glu	Leu	Gln	Lys	Ile 670	Leu	Glu
Leu	Lys	Ala 675	Gly	Ile	Ser	Asn	Lys 680	Ser	Asn	Arg	Tyr	Asn 685	Asp	Asn	Tyr
Asn	Asn 690	Tyr	Ile	Ser		Сув 695	Ser	Ile	Ile	Thr	Авр 700	Leu	Ser	Lys	Phe
Asn 705	Gln	Ala	Phe		Tyr 710	Glu	Thr	Ser	Сув	Ile 715	Сув	Ser	Asp	Val	Leu 720
				725					730		Trp			735	
le	Pro		Val 740	Thr	Ile :	Ile		Thr 745	Tyr	Arg	нів		Pro 750	Pro	Phe

- 396 -

Ile	Lys	<b>Asp</b> 755	His	Val	Val	Asn	<b>Leu</b> 760	Asn	Glu	Val	Asp	Glu 765	Gln	Ser	Gl
Leu	<b>Tyr</b> 770	Arg	Tyr	His	Met	Gly 775	Gly	Ile	Glu	Gly	Trp 780	Сув	Gln	Lys	Let
Trp 785	Thr	Ile	Glu	Ala	Ile 790	Ser	Leu	Leu	Asp	Leu 795	Ile	Ser	Leu	Lys	Gl <sub>3</sub> 800
Lys	Phe	Ser	Ile	Thr 805	Ala	Leu	Ile	Asn	Gly 810	Asp	Asn	Gln	Ser	Ile 815	Aag
Ile	Ser	Lys	Pro 820	Val	Arg	Leu	Ile	Glu 825	Gly	Gln	Thr	His	Ala 830	Gln	Ala
Asp	Tyr	<b>Leu</b> 835	Leu	Ala	Leu	Asn	Ser 840	Leu	Lys	Leu	Leu	Tyr 845	Lys	Glu	Туг
Ala	Gly 850	Ile	Gly	His	Lys	Leu 855	Lys	Gly	Thr	Glu	Thr 860	Tyr	Ile	Ser	Arg
Asp 865	Met	Gln	Phe	Met	Ser 870	Lys	Thr	Ile	Gln	His 875	Asn	Gly	Val	Tyr	Ту: 880
Pro	Ala	Ser	Ile	Lys 885	Lys	Val	Leu	Arg	Val 890	Gly	Pro	Trp	Ile	Asn 895	Thr
Ile	Leu	Asp	Asp 900	Phe	Lys	Val	Ser	Leu 905	Glu	Ser	Ile	Gly	Ser 910	Leu	Thr
		Leu 915					920					925			
Arg	Asn 930	Ile	Trp	Leu	Tyr	Asn 935	Gln	Ile	Ala	Leu	Gln 940	Leu	Arg	Asn	Kis
Ala 945	Leu	Сув	Asn	Asn	<b>Lys</b> 950	Leu	Tyr	Leu	qaA	Ile 955	Leu	Lys	Val	Leu	P60
His	Leu	Lys	Thr	Phe 965	Phe	Asn	Leu	qaA	Ser 970	Ile	Asp	Met	Ala	Leu 975	
Leu ·	Tyr	Met	<b>Asn</b> 980	Leu	Pro	Met	Leu	Phe 985	Gly	Gly	Gly	Asp	Pro 990	Asn	Leu
Leu	Tyr	Arg 995	Ser	Phe	Tyr	Arg	Arg 1000		Pro	Asp	Phe	Leu 1005		Glu	Ala
Ile	Val 1010	His	Ser	Val	Phe	Val 1015		Ser	Tyr	Tyr	Thr 1020		His	Asp	Leu

- 397 -

Gln	Asp	Lys	Leu	Gln	Asp	Leu	Pro	Asp	Asp	Arg	Leu	Asn	Lys	Phe	Leu
1025	į				1030	)				103	5				1040

- Thr Cys Val Ile Thr Phe Asp Lys Asn Pro Asn Ala Glu Phe Val Thr
  1045 1050 1055
- Leu Met Arg Asp Pro Gln Ala Leu Gly Ser Glu Arg Gln Ala Lys Ile 1060 1065 1070
- Thr Ser Glu Ile Asn Arg Leu Ala Val Thr Glu Val Leu Ser Ile Ala 1075 1080 1085
- Pro Asn Lys Ile Phe Ser Lys Ser Ala Gln His Tyr Thr Thr Glu 1090 1095 1100
- Ile Asp Leu Asn Asp Ile Met Gln Asn Ile Glu Pro Thr Tyr Pro His 1105 1110 1115 1120
- Gly Leu Arg Val Val Tyr Glu Ser Leu Pro Phe Tyr Lys Ala Glu Lys
  1125 1130 1135
- Ile Val Asn Leu Ile Ser Gly Thr Lys Ser Ile Thr Asn Ile Leu Glu 1140 1145 1150
- Lys Thr Ser Ala Ile Asp Thr Thr Asp Ile Asn Arg Ala Thr Asp Met 1155 1160 1165
- Met Arg Lys Asn Ile Thr Leu Leu Ile Arg Ile Leu Pro Leu Asp Cys 1170 1175 1180
- Asn Lys Asp Lys Arg Glu Leu Leu Ser Leu Glu Asn Leu Ser Ile Thr 1185 1190 1195 1200
- Glu Leu Ser Lys Tyr Val Arg Glu Arg Ser Trp Ser Leu Ser Asn Ile 1205 1210 1215
- Val Gly Val Thr Ser Pro Ser Ile Met Phe Thr Met Asp Ile Lys Tyr 1220 1225 1230
- Thr Thr Ser Thr Ile Ala Ser Gly Ile Ile Ile Glu Lys Tyr Asn Val 1235 1240 1245
- Asn Ser Leu Thr Arg Gly Glu Arg Gly Pro Thr Lys Pro Trp Val Gly 1250 1260
- Ser Ser Thr Gln Glu Lys Lys Thr Met Pro Val Tyr Asn Arg Gln Val 1265 1270 1275 1280
- Leu Thr Lys Lys Gln Arg Asp Gln Ile Asp Leu Leu Ala Lys Leu Asp 1295
- Trp Val Tyr Ala Ser Ile Asp Asn Lys Asp Glu Phe Met Glu Glu Leu

- 398 -

			1300	)				130	5				1310	)	
Ser	Thr	Gly 1315		Leu	Gly	Leu	Ser 1320		Glu	Lys	Ala	Lys 1325		Leu	Phe
Pro	Gln 1330		Leu	Ser	Val	Asn 1335		Leu	His	Arg	Leu 1340		Val	Ser	Ser
Arg 1345		Сув	Glu	Phe	Pro 1350		Ser	Ile	Pro	Ala 1355	_	Arg	Thr	Thr	<b>Asn</b> 1360
Tyr	His	Phe	Asp	Thr 1365	Ser	Pro	Ile	Asn.	His 1370		Leu	Thr	Glu	Lys 1375	-
Gly	Авр	Glu	Asp 1380		Авр	Ile	Val	Phe 1385		Asn	Сув	Ile	Ser 1390		Gly
Leu	Ser	Leu 1395		Ser	Val	Val	Glu 1400		Phe	Thr	Asn	Ile 1405	_	Pro	Asn
Arg	Ile 1410		Leu	Ile	Pro	Lys 1419		Asn	Glu	Ile	His 1420		Met	Lys	Pro
Pro 1425		Phe	Thr	Gly	Asp 1430		Авр	Ile	Ile	Lys 1435		Lys	Gln	Val	Ile 1440
Gln	Lva	aln	Wi o	W	<b>5</b> L -		_	_	_		_	_			
	-,-	<b>J.</b>	UIB	1445		Leu	Pro	Авр	Lys 1450		Ser	Leu	Thr	Gln 1455	-
Val				1445 Leu	5				1450 Leu	)				1455 His	
Val	Glu	Leu	Phe 1460 Leu	1445 Leu	5	Asn	Lys	Ala 1465 Lys	1450 Leu	Lys	Ser	Gly	Ser 1470 Phe	1455 His	Ile
Val Asn	Glu Ser	Leu Asn 1475	Phe 1460 Leu	1445 Leu	Ser	Asn Val	Lys His 1480 Leu	Ala 1465 Lys	1450 Leu ; Met	Lys Ser	Ser	Gly Tyr 1485	Ser 147( Phe	1455 His ) His	Ile Asn
Val Asn Ala	Glu Ser Tyr 1490 Leu	Leu Asn 1475	Phe 1460 Leu Leu	Leu Ile Ser	Ser Leu	Asn Val Asn 1495 Lys	Lys His 1480 Leu	Ala 1465 Lys Ala	1450 Leu ; Met Gly	Lys Ser His	Ser Asp Trp 1500	Gly Tyr 1485	Ser 1470 Phe Leu	His His His	Ile Asn Ile
Val Asn Ala Gln	Glu Ser Tyr 1490 Leu	Leu Asn 1475 Ile Met	Phe 1460 Leu Leu	Leu Ile Ser	Ser Leu Thr Ser 1510	Asn Val Asn 1495 Lys	Lys His 1480 Leu Gly	Ala 1465 Lys Ala	Leu Het Gly	Lys Ser His Glu 1515	Asp Trp 1500 Lys	Gly Tyr 1485 Ile	Ser 1470 Phe Leu	His His Ile	Ile Asn Ile Glu 1520 Asn
Val Asn Ala Gln 1505	Glu Ser Tyr 1490 Leu	Leu Asn 1475 Ile Met	Phe 1460 Leu Leu Lys	1445 Leu Ile Ser Asp 1525	Ser Leu Thr Ser 1510	Asn Val Asn 1495 Lys	Lys His 1480 Leu Gly	Ala 1465 Lys Ala Ile	Leu  Met  Gly  Phe  Asn 1530	Lys Ser His Glu 1515	Asp Trp 1500 Lys	Tyr 1485 Ile Asp	Ser 1470 Phe Leu Trp	His His Ile Gly Phe 1535	Ile Asn Ile Glu 1520 Asn
Val Asn Ala Gln 1505 Gly Ala	Glu Ser Tyr 1490 Leu Tyr	Leu Asn 1475 Ile Met Ile	Phe 1460 Leu Leu Lys Thr Thr 1540	Leu Ile Ser Asp 1525	Ser Leu Thr Ser 1510	Asn Val Asn 1495 Lys Met	Lys His 1480 Leu Gly Phe	Ala 1465 Lys Ala Ile Phe 1545 Ser	Leu  Met  Gly  Phe  Asn 1530	Lys Ser His Glu 1515 Leu	Asp Trp 1500 Lys Asn	Tyr 1485 Ile Asp Val	Ser 1470 Phe Leu Trp Phe Gly 1550	His His Gly Phe 1535	Ile Asn Ile Glu 1520 Asn

- Gln Lys Val Ile Lys Tyr Ile Val Asn Gln Asp Thr Ser Leu Arg Arg 1585 1590 1595 1600
- Ile Lys Gly Cys His Ser Phe Lys Leu Trp Phe Leu Lys Arg Leu Asp 1605 1610 1615
- Asn Ala Lys Phe Thr Val Cys Pro Trp Val Val Asn Ile Asp Tyr His 1620 · 1625 1630
- Pro Thr His Met Lys Ala Ile Leu Ser Tyr Ile Asp Leu Val Arg Met 1635 1640 1645
- Gly Leu Ile Asn Val Asp Lys Leu Thr Ile Lys Asn Lys Asn Lys Phe 1650 1655 1660
- Asn Asp Glu Phe Tyr Thr Ser Asn Leu Phe Tyr Ile Ser Tyr Asn Phe 1665 1670 1675 ~1680
- Ser Asp Asn Thr His Leu Leu Thr Lys Gln Ile Arg Ile Ala Asn Ser 1685 1690 1695
- Glu Leu Glu Asp Asn Tyr Asn Lys Leu Tyr His Pro Thr Pro Glu Thr 1700 1705 1710
- Leu Glu Asn Met Ser Leu Ile Pro Val Lys Ser Asn Asn Ser Asn Lys 1715 1720 1725
- Pro Lys Phe Cys Ile Ser Gly Asn Thr Glu Ser Met Met Ser Thr 1730 1740
- Phe Ser Ser Lys Met His Ile Lys Ser Ser Thr Val Thr Thr Arg Phe 1745 1750 1755 1766
- Asn Tyr Ser Lys Gln Asp Leu Tyr Asn Leu Phe Pro Ile Val Val Ile 1765 1770 1775
- Asp Lys Ile Ile Asp His Ser Gly Asn Thr Ala Lys Ser Asn Gln Leu 1780 1785 1790
- Tyr Thr Thr Ser His Gln Thr Ser Leu Val Arg Asn Ser Ala Ser 1795 1800 1805
- Leu Tyr Cys Met Leu Pro Trp His His Val Asn Arg Phe Asn Phe Val 1810 1815 1820
- Phe Ser Ser Thr Gly Cys Lys Ile Ser Ile Glu Tyr Ile Leu Lys Asp 1825 1830 1835 1840
- Leu Lys Ile Lys Asp Pro Ser Cys Ile Ala Phe Ile Gly Glu Gly Ala 1845 1850 1855

- Gly Asn Leu Leu Leu Arg Thr Val Val Glu Leu His Pro Asp Ile Arg 1860 1865 1870
- Tyr Ile Tyr Arg Ser Leu Lys Asp Cys Asn Asp His Ser Leu Pro Ile 1875 1880 1885
- Glu Phe Leu Arg Leu Tyr Asn Gly His Ile Asn Ile Asp Tyr Gly Glu 1890 1895 1900
- Asn Leu Thr Ile Pro Ala Thr Asp Ala Thr Asn Asn Ile His Trp Ser 1905 1910 1915 1920
- Tyr Leu His Ile Lys Phe Ala Glu Pro Ile Ser Ile Phe Val Cys Asp 1925 1930 1935
- Ala Glu Leu Pro Val Thr Ala Asn Trp Ser Lys Ile Ile Glu Trp 1940 1945 1950
- Ser Lys His Val Arg Lys Cys Lys Tyr Cys Ser Ser Val Asn Arg Cys 1955 1960 1965
- Ile Leu Ile Ala Lys Tyr His Ala Gln Asp Asp Ile Asp Phe Lys Leu 1970 1975 1980
- Asp Asn Ile Thr Ile Leu Lys Thr Tyr Val Cys Leu Gly Ser Lys Leu 1985 1990 1995 2000
- Lys Gly Ser Glu Val Tyr Leu Ile Leu Thr Ile Gly Pro Ala Asn Ile 2005 2010 2015
- Leu Pro Val Phe Asp Val Val Gln Asn Ala Lys Leu Ile Leu Ser Arg 2020 2025 2030
- Thr Lys Asn Phe Ile Met Pro Lys Lys Thr Asp Lys Glu Ser Ile Asp 2035 2040 2045
- Ala Val Ile Lys Ser Leu Ile Pro Phe Leu Cys Tyr Pro Ile Thr Lys 2050 2055 2060
- Lys Gly Ile Lys Thr Ser Leu Ser Lys Leu Lys Ser Val Val Asn Gly 2065 2070 2075 2080
- Asp Ile Leu Ser Tyr Ser Ile Ala Gly Arg Asn Glu Val Phe Ser Asn 2085 2090 2095
- Lys Leu Ile Asn His Lys His Met Asn Ile Leu Lys Trp Leu Asp His 2100 2105 2110
- Val Leu Asn Phe Arg Ser Ala Glu Leu Asn Tyr Asn His Leu Tyr Met 2115 2120 2125
- Ile Glu Ser Thr Tyr Pro Tyr Leu Ser Glu Leu Leu Asn Ser Leu Thr

- 401 -

2130 2135 2140 Thr Asn Glu Leu Lys Lys Leu Ile Lys Ile Thr Gly Ser Val Leu Tyr 2150 2155 Asn Leu Pro Asn Glu Gln 2165 (2) INFORMATION FOR SEQ ID NO:35: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: RNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35: CATATCACTC ACTCTGGGAT GGAG 24 (2) INFORMATION FOR SEQ ID NO:36: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: RNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36: TCAGAACATC AAGCACCGCC 20 (2) INFORMATION FOR SEQ ID NO:37: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

- 402 -

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
ACAG'	TCAAGA CTGAGATGAG	20
(2)	INFORMATION FOR SEQ ID NO:38:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
AAGA	GTCAGA TACATGTGGA	20
(2)	INFORMATION FOR SEQ ID NO:39:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
ACAT	GAATCA GCCTAAAGTC	20
(2)	INFORMATION FOR SEQ ID NO:40:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 25 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: RNA (genomic)	
	/** MATHEARD TTER. WIN (Advance)	

- 403 -

	(x1) SEQUENCE DESCRIPTION: SEQ ID NO:40:	
CCG	AAAGAGT TCCTGCGTTA CGACC	25
(2)	INFORMATION FOR SEQ ID NO:41:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
CAG	TCCACAC AAGTACCAGG	20
(2)	INFORMATION FOR SEQ ID NO:42:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
GTCA	AGAAGCT GTGGACCATC	20
(2)	INFORMATION FOR SEQ ID NO:43:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	

- 404 -

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
AATA	TTGCTA CAACAATGGC	20
(2)	INFORMATION FOR SEQ ID NO:44:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:	
		_
ACTC	TTCATT CCTAGACTGG	20
(2)	INFORMATION FOR SEQ ID NO:45:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
GTCC	CANTTAT GACTATGAAC	20
(2)	INFORMATION FOR SEQ ID NO:46:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	

- 405 -

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
AGAI	ACAGACA TGAAGCTTGC	20
(2)	INFORMATION FOR SEQ ID NO:47:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
CCAA	ACAAGGA ATGCTTCTAG	20
(2)	INFORMATION FOR SEQ ID NO:48:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 25 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
ACAG	GCACTAT CTATGATTGA CCTGG	25
(2)	INFORMATION FOR SEQ ID NO:49:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: RNA (genomic)	
	,	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

- 406 -

GCAACATGGT TTACACATGC	20
(2) INFORMATION FOR SEQ ID NO:50:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
AGATTGAGAG TTGATCCAGG	20
(2) INFORMATION FOR SEQ ID NO:51:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:	
AGGAGATACT TAAACTAAGC	20
(2) INFORMATION FOR SEQ ID NO:52:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:	

- 407 -

20

TAAGCTTATG CCTTTCAGCG

(2)	INFORMATION FOR SEQ ID NO:53:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:	
TTAA	CGGACC TAAGCTGTGC	20
(2)	INFORMATION FOR SEQ ID NO:54:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:	
GAAA	CAGATT ATTATGACGG	20
(2)	INFORMATION FOR SEQ ID NO:55:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 24 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:	
ceca	CTATCT AGGTGAACTT CAGG	

- 408 -

(2) INFORMATION FOR SEQ ID NO:56:

		4
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:	
ATTT	GGATAT GGAATATGAG	20
(2)	INFORMATION FOR SEQ ID NO:57:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:	
ACTC	AACTGA ACTACCAGTG	20
(2)	INFORMATION FOR SEQ ID NO:58:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:	
AAGAJ	ACATCA TGTATTTCAG	20

20

- 409 -

(2) INFORMATION FOR SEQ ID NO:59:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:	
TTATCAACGC ACTGCTCATG	20
(2) INFORMATION FOR SEQ ID NO:60:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 25 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:	
ATTTTCAGCA ATCACTTGGC ATGCC	25
(2) INFORMATION FOR SEQ ID NO:61:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 20 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:	
GCCTCTGTGC AAACAAGCTG	20
(2) INFORMATION FOR SEQ ID NO:62:	

- 410 -

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:	
TCT	CTAGTTA CTCTAGCAGC	20
(2)	INFORMATION FOR SEQ ID NO:63:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:	
AGGT	ICGTTGT TTGTGAGGAG	20
(2)	INFORMATION FOR SEQ ID NO:64:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO:64:	
	· ·	
TCGI	CCTCTT CTTTACTGTC	20
(2)	INFORMATION FOR SEQ ID NO:65:	
. – ,	FOU DEE 4D MAIANI	

- 411 -

(i) SEQUENCE CHARACTERISTICS:

	(A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:	
CCGT	CCTCGA GCTAGCCTCG	20
(2)	INFORMATION FOR SEQ ID NO:66:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:	20
(2)	INFORMATION FOR SEQ ID NO:67:	
	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
I	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:	•
GGGTT	IGGTAC ATAGCTCTGC	20
(2) 1	INFORMATION FOR SEQ ID NO:68:	
	(i) CENTENCE CURDACTEDICTICS.	

- 412 -

	(m, mandin. 25 base pails	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	·	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:	
CAC	CCATCTG ATATTTCCCT GATGG	25
		-
(2)	INFORMATION FOR SEQ ID NO:69:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:	
TGG	TTGACAG TACAAATCTG	20
(2)	INFORMATION FOR SEQ ID NO:70:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: RNA (genomic)	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:	
CTG	LAATGGG AAGATTGTGC	20
(2)	INFORMATION FOR SEQ ID NO:71:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 20 base pairs	

- 413 -

		(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	( <b>ii</b> )	MOLECULE TYPE: RNA (genomic)	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:71:	
AGC	AATCT	AC ACTGCCTACC	20
(2)	INFO	RMATION FOR SEQ ID NO:72:	
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: RNA (genomic)	
		SEQUENCE DESCRIPTION: SEQ ID NO:72:	
TCAC	AGAT	GA TTCAATTATC	20
(2)	INFO	RMATION FOR SEQ ID NO:73:	
	(1)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 20 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: RNA (genomic)	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:73:	
GATC	CTAG	AT ATAAGTTCTC	20
(2)	INFO	RMATION FOR SEQ ID NO:74:	
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 21 base pairs  (B) TYPE: pucleic acid	

- 414 -

(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:	
ACCAAACAAA GTTGGGTAAG G	21
(2) INFORMATION FOR SEQ ID NO:75:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 32 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:	
GGGGGATCCA TCCCTAATCC TGCTCTTGTC CC	32
(2) INFORMATION FOR SEQ ID NO:76:	
41) 4-4	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
4111	
(ii) MOLECULE TYPE: RNA (genomic)	
(wi) GRAMMAT DECONTRACTOR OF THE SECOND	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:	
GATTCCTCTG ATGGCTCCAC	20
(2) INFORMATION FOR SEQ ID NO:77:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS, single	

- 415 -

(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:	
TAACAGTCAA GGAGACCAAA G	21
(2) INFORMATION FOR SEQ ID NO:78:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 32 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:78:	
GGGAAGCTTA ACCCTAATCC TGCCCTAGGT GG	32
(2) INFORMATION FOR SEQ ID NO:79:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 22 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: RNA (genomic)	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:	
ACCAGACAAA GCTGGGAATA GA	22

- 416 -

## What is claimed is:

- 1. An isolated, recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene.
- 2. The virus of Claim 1 wherein the virus is from the Family Paramyxoviridae.
- 3. The virus of Claim 2 wherein the virus is from the Subfamily Paramyxovirinae.
- 4. The virus of Claim 3 wherein the virus is from the Genus Morbillivirus.
- 5. The virus of Claim 4 wherein the virus is measles virus.
  - 6. The measles virus of Claim 5 wherein:
  - (a) the at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 26 (A → T), nucleotide 42 (A → T or A → C) and nucleotide 96 (G → A), where these nucleotides are presented in positive strand, antigenomic, message sense; and
  - (b) the at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 331 (isoleucine → threonine), 1409 (alanine → threonine), 1624 (threonine → alanine), 1649 (arginine → methionine), 1717

(aspartic acid  $\rightarrow$  alanine), 1936 (histidine  $\rightarrow$  tyrosine), 2074 (glutamine  $\rightarrow$  arginine) and 2114 (arginine  $\rightarrow$  lysine).

- 7. The virus of Claim 3 wherein the virus is from the Genus Paramyxovirus.
- 8. The virus of Claim 7 wherein the virus is human parainfluenzae virus type 3 (PIV-3).
  - 9. The PIV-3 of Claim 8 wherein:
  - (a) the at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 23 (T → C), nucleotide 24 (C → T), nucleotide 28 (G → T) and nucleotide 45 (T → A), where these nucleotides are presented in positive strand, antigenomic, message sense; and
  - (b) the at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 942 (tyrosine → histidine), 992 (leucine → phenylalanine), 1292 (leucine → phenylalanine), and 1558 (threonine → isoleucine).
- 10. The virus of Claim 3 wherein the virus is from the Genus Rubulavirus.
- 11. The virus of Claim 2 wherein the virus is from the Subfamily Pneumovirinae.
- 12. The virus of Claim 11 wherein the virus is from the Genus Pneumovirus.

- 418 -

- 13. The virus of Claim 12 wherein the virus is human respiratory syncytial virus (RSV) subgroup B.
  - 14. The virus of Claim 13 wherein:
  - (a) the at least one attenuating mutation in the 3' genomic promoter region is selected from the group consisting of nucleotide 4 (C → G) and the insertion of an additional A in the stretch of A's at nucleotides 6-11, where these nucleotides are presented in positive strand, antigenomic, message sense; and
  - (b) the at least one attenuating mutation in the RNA polymerase gene is selected from the group consisting of nucleotide changes which produce changes in an amino acid selected from the group consisting of residues 353 (arginine → lysine), 451 (lysine → arginine), 1229 (aspartic acid → asparagine), 2029 (threonine → isoleucine) and 2050 (asparagine → aspartic acid).
- 15. The virus of Claim 1 wherein the virus is from the Family Rhabdoviridae.
- 16. The virus of Claim 1 wherein the virus is from the Family Filoviridae.
- 17. A vaccine comprising an isolated, recombinantly-generated, attenuated, nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales according to Claim 1 and a physiologically acceptable carrier.
- 18. The vaccine of Claim 17 comprising a measles virus according to Claim 5 and a physiologically acceptable carrier.

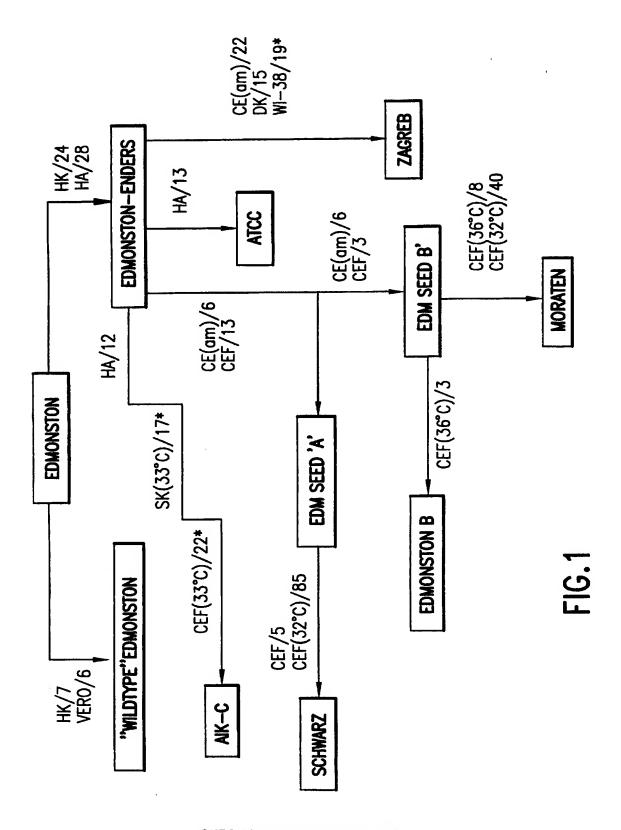
- 19. The vaccine of Claim 18 comprising a measles virus according to Claim 6 and a physiologically acceptable carrier.
- 20. The vaccine of Claim 17 comprising a PIV-3 according to Claim 8 and a physiologically acceptable carrier.
- 21. The vaccine of Claim 20 comprising a PIV-3 according to Claim 9 and a physiologically acceptable carrier.
- 22. The vaccine of Claim 17 comprising an RSV subgroup B according to Claim 13 and a physiologically acceptable carrier.
- 23. The vaccine of Claim 22 comprising an RSV subgroup B according to Claim 14 and a physiologically acceptable carrier.
- 24. A method for immunizing an individual to induce protection against a nonsegmented, negativesense, single stranded RNA virus of the Order Mononegavirales which comprises administering to the individual the vaccine of Claim 17.
- 25. The method of Claim 24 wherein the vaccine is the vaccine of Claim 18.
- 26. The method of Claim 25 wherein the vaccine is the vaccine of Claim 19.
- 27. The method of Claim 24 wherein the vaccine is the vaccine of Claim 20.
- 28. The method of Claim 27 wherein the vaccine is the vaccine of Claim 21.
- 29. The method of Claim 24 wherein the vaccine is the vaccine of Claim 22.
- 30. The method of Claim 29 wherein the vaccine is the vaccine of Claim 23.

- 31. An isolated nucleic acid molecule comprising a measles virus sequence in positive strand, antigenomic message sense selected from the group consisting of 1977 wild-type strain (SEQ ID NO:3), 1983 wild-type strain (SEQ ID NO:5) where the nucleotide 2499 is G or C, Montefiore wild-type strain (SEQ ID NO:7), Rubeovax™ vaccine strain (SEQ ID NO:9), where the nucleotide 2143 is T or C, Moraten vaccine strain (SEQ ID NO:11), where the nucleotide 4917 is C and the nucleotide 4924 is C, and Zagreb vaccine strain (SEQ ID NO:13), and the complementary genomic sequences thereof.
- 32. An isolated nucleic acid molecule comprising a PIV-3 sequence in positive strand, antigenomic message sense selected from the group consisting of cp45 vaccine strain grown in fetal rhesus lung cells (SEQ ID NO:19) and cp45 vaccine strain grown in Vero cells (SEQ ID NO:21), and the complementary genomic sequences thereof.
- transcription vector comprising an isolated nucleic acid molecule encoding a genome or antigenome of a nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales having at least one attenuating mutation in the 3' genomic promoter region and having at least one attenuating mutation in the RNA polymerase gene, together with at least one expression vector which comprises at least one isolated nucleic acid molecule encoding the trans-acting proteins necessary for encapsidation, transcription and replication, whereby upon expression an infectious attenuated virus is produced.

- 34. The composition of Claim 33 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes a measles virus according to Claim 5 and the at least one expression vector comprises at least one isolated nucleic acid molecule encoding the trans-acting proteins N, P and L.
- 35. The composition of Claim 34 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes a measles virus according to Claim 6.
- 36. The composition of Claim 33 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes a PIV-3 according to Claim 8 and the at least one expression vector comprises at least one isolated nucleic acid molecule encoding the transacting proteins NP, P and L.
- 37. The composition of Claim 36 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes a PIV-3 according to Claim 9.
- 38. The composition of Claim 33 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes an RSV subgroup B according to Claim 13 and the at least one expression vector comprises at least one isolated nucleic acid molecule encoding the trans-acting proteins N, P, L and M2.
- 39. The composition of Claim 38 wherein the transcription vector comprises an isolated nucleic acid molecule which encodes an RSV subgroup B according to Claim 14.
- 40. A method for producing infectious attenuated nonsegmented, negative-sense, single stranded RNA virus of the Order Mononegavirales which comprises transforming or transfecting host cells with

the at least two vectors of Claim 33 and culturing the host cells under conditions which permit the co-expression of these vectors so as to produce the infectious attenuated virus.

- 41. The method of Claim 40 wherein the virus is the measles virus of Claim 5.
- 42. The method of Claim 41 wherein the virus is the measles virus of Claim 6.
- 43. The method of Claim 40 wherein the virus is the PIV-3 of Claim 8.
- 44. The method of Claim 43 wherein the virus is the PIV-3 of Claim 9.
- 45. The method of Claim 40 wherein the virus is the RSV subgroup B of Claim 13.
- 46. The method of Claim 45 wherein the virus is the RSV subgroup B of Claim 14.



SUBSTITUTE SHEET (RULE 26)

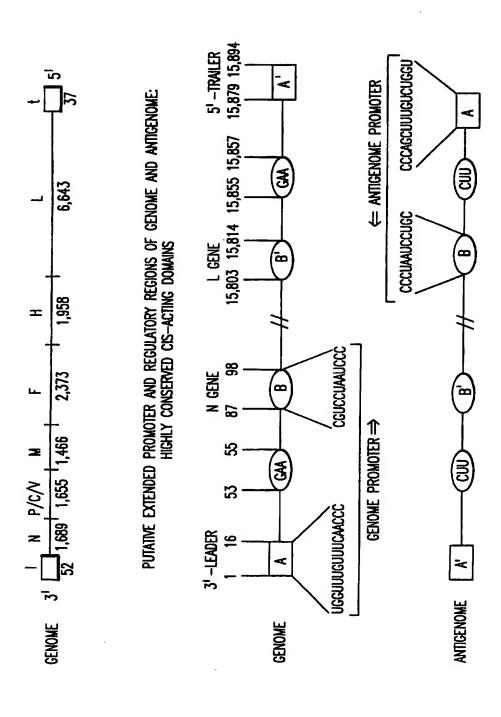
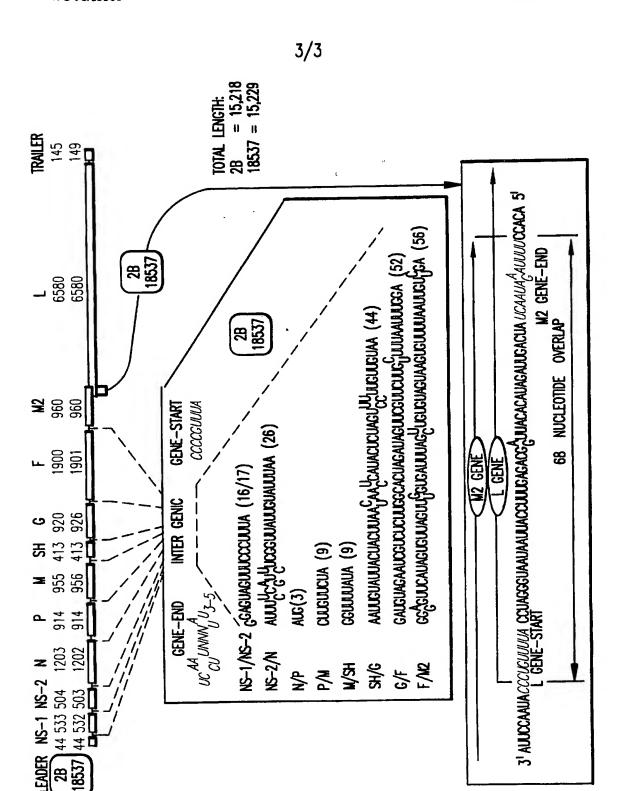


FIG.2



SUBSTITUTE SHEET (RULE 26)